



PROJECT MUSE®

Formal grammar, usage probabilities, and auxiliary contraction

Joan Bresnan

Language, Volume 97, Number 1, March 2021, pp. 108-150 (Article)

Published by Linguistic Society of America

DOI: <https://doi.org/10.1353/lan.2021.0003>

A thumbnail image of the table of contents for the journal 'LANGUAGE'. The title 'LANGUAGE' is at the top, followed by 'A JOURNAL OF THE LINGUISTIC SOCIETY OF AMERICA'. Below this, there is a list of articles with their authors and page numbers. The text is small and difficult to read, but it appears to be a standard table of contents layout.

➔ *For additional information about this article*

<https://muse.jhu.edu/article/785541>

FORMAL GRAMMAR, USAGE PROBABILITIES, AND AUXILIARY CONTRACTION

JOAN BRESNAN

Stanford University

This article uses formal and usage-based data and methods to argue for a hybrid model of English tensed auxiliary contraction combining lexical syntax with a dynamic exemplar lexicon. The hybrid model can explain why the contractions involve lexically specific phonetic fusions that have become morphologized and lexically stored, yet remain syntactically independent, and why the probability of contraction itself is a function of the adjacent cooccurrences of the subject and auxiliary in usage, yet is also subject to the constraints of the grammatical context. Novel evidence includes a corpus study and a formal analysis of a multiword expression of classic usage-based grammar.*

Keywords: contraction, clitic, corpus, probability, LFG, exemplar, usage

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 2001, 2002, 2006 at the level of word phonetics). 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.

The present study presents evidence for a proposed hybrid model at the syntactic level for English tensed auxiliary contractions, using LEXICAL-FUNCTIONAL GRAMMAR (LFG) with lexical sharing (Wescoat 2002, 2005) as the representational basis for the syntax, and a dynamic exemplar model of the mental lexicon similar to the hybrid model proposals at the phonetic word level along the lines of Pierrehumbert 2001. However, the aim of this study is not to present a formalization of a particular hybrid model or to argue for a specific formal grammar. The aim is to show the empirical and theoretical value of combining formal and usage-based data and methods into a shared framework—a theory of lexical syntax and a dynamic usage-based lexicon that includes multiword sequences.

* I am grateful to several colleagues who helped me advance this study: to Jen Hay for making data for the corpus study available while I had a Strategic Fellowship at the New Zealand Institute for Language, Brain, and Behaviour in 2015; to Arto Anttila for his collaborations with me in corpus phonology and syntax from our joint Stanford course in 2012 and through summer projects in 2015 and 2017; to Adams Bodomo for his invitation and very generous support of my presentation at LFG '18 at the University of Vienna; to Avery Andrews, Joan Bybee, and Janet Pierrehumbert in particular among those who gave me comments on the first draft; and to the editorial team of this journal, including three anonymous referees who gave detailed and helpful comments on subsequent drafts.

Tensed auxiliary¹ contractions in English are particularly interesting because the contracting elements appear to cross the boundary between the major constituents of the sentence, namely the subject and the verb phrase, as in *You're sick*: here *are* serves as the main verb of the sentence and contracts with the subject *you*. These contractions are not semantic constituents of their larger utterances. 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 bloody hell's knocking* (from the Canterbury Corpus; Gordon et al. 2004) are not compositional components in the semantics of the sentences. Nor are they syntactic constituents: witness **Who's do you think coming?* vs. *Who do you think is coming?* (Anderson 2008:174), or **It's you're that sick* vs. *It's you that are sick*.²

Nevertheless, tensed auxiliary contractions in some contexts show signs of being units. For example, the clitic auxiliary *'s* provides a coda of the open syllables of *law* and *hell*, which select the voiced variant [z] (in contrast to *his wife's a teacher*, which selects [s]). For these reasons tensed auxiliary contraction has long been treated in formal linguistic frameworks as SIMPLE CLITICIZATION (Zwicky 1977), a phonological grouping of two adjacent nonconstituent words belonging to the surface syntactic phrasings of metrical and prosodic phonology (Selkirk 1984, 1996, Inkelas & Zec 1993, Anderson 2008, Anttila 2017, Ito & Mester 2018)—purely supralephical phonological processes.

Yet that is far from the whole story: a number of researchers have pointed out morphophonological properties of the most common auxiliary contractions that are signs of the contracted forms being lexically stored (Kaisse 1985, A. Spencer 1991, Bybee & Scheibman 1999, Scheibman 2000, Wescot 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, Bybee 2002, Frank & Jaeger 2008, Bresnan & Spencer 2012, J. Spencer 2014, Barth & Kapatsinski 2017, Barth 2019).

What appears to be needed in order to explain fully the properties of tensed auxiliary contractions is a theory of their representations that simultaneously accounts for their syntactic nonconstituency and adjacency constraints, their lexical morphophonology, and 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,

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 sleeps n't*; and *Is she sleeping/sleepy?* vs. **Sleeps she?*. See Huddleston & Pullum 2002 for discussion of a fuller set of distinguishing properties of auxiliary verbs.

² These interrogative and clefting constructions otherwise allow larger syntactic constituents to appear with the focused phrase, as in *At what time do you think she's coming?*, *It's with Louise that she was running*.

Zwicky & Pullum 1983, Selkirk 1984, 1996, Klavans 1985, Inkelas & Zec 1993, Wilder 1997, Barron 1998, Sadler 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 information-theoretic measures (for example, Frank & Jaeger 2008, Barth 2011, 2019, J. Spencer 2014, Barth & Kapatsinski 2017). 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, Rickford et al. 1991, McElhinny 1993, MacKenzie 2012, 2013).

The present study of tensed auxiliary contraction proposes that the formal syntactic theory of lexical sharing in LFG, combined with a hybrid exemplar-dynamic model of the mental lexicon, can provide the necessary combined approach. Lexical sharing in LFG was originally designed to account for narrowly defined types of cases where lexical units do not match constituent structure units, such as contractions of preposition-determiner combinations (for example, German *zum*, *am*, *im*, *ins* and French *du*, *au*, *des*, *aux*, discussed by Bybee 2002 and Wescoat 2007, among others) and contractions of simple clitics, like English tensed auxiliary contractions (also discussed by Bybee 2002, 2010 and Wescoat 2005, among others). However, as the present study shows, lexical sharing naturally extends to the lexicalization of multiword sequences in larger constructions.³ While the formal analyses by themselves provide insights into the grammar of tensed auxiliary contraction, they ignore the explanatory role of usage probabilities in syntactic lexicalizations. But usage-based linguistic studies of tensed auxiliary contraction have seldom presented fully articulated proposals for their syntactic representations, and without adequate representations of the grammatical context they cannot account for constraints on contraction. The present study therefore contributes to both formal and usage-based lines of research.

The first three sections below outline some of the main findings of usage-based linguistics on tensed auxiliary contractions and show how they are explained theoretically (§§1–3). The following three sections (§§4–6) outline the main findings of formal research on tensed auxiliary contraction, and show how they are captured in the particular formal framework of lexical syntax adopted here. The next section (§7) presents a hybrid model that synthesizes the formal and usage-based findings, and the following sections present novel evidence for such a hybrid (§§8–10): a corpus study of *is* contraction, a formal analysis of gradient subtypes of contracting auxiliaries, and the extension of the formal grammar of auxiliary contraction to a multiword expression of classic usage-based grammar (Bybee & Scheibman 1999) that brings out surprising parallels with tensed auxiliary contraction. The concluding section compares several alternative frameworks (§11).

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

³ Although most previous work with lexical sharing in LFG has concerned contraction, cliticization, and portmanteau-word phenomena with prepositions and determiners (e.g. Wescoat 2007, 2009, Broadwell 2008, Alsina 2010, Lowe 2016), Broadwell 2007 extends the theory to certain multiword expressions that form phonological words in Zapotec.

linguistics. The data consist of constructed examples, Web examples, and corpus examples (with quantitative analysis of the latter). If an example is not labeled ‘authentic’, it is constructed. The primary sources of authentic data are the Buckeye Corpus (BC) (Pitt et al. 2005) of spoken mid-American English and the Canterbury Corpus (CC) (Gordon et al. 2004) of spoken New Zealand English.⁴

Judgments of constructed examples are validated or corrected, where possible, using examples from corpora as well as MacKenzie’s (2012) careful corpus study of auxiliary contraction in spoken English, and finally the Web, for a few rarer constructions. But sometimes judgments simply represent ‘working evidence’ to motivate a hypothesis until substantiating data can be obtained.

1. USAGE AND PHONETIC REDUCTION. A major finding of usage-based linguistics is that more probable words and multiword expressions are phonetically more reduced and become lexically stored (Bybee 2001, 2006, Bybee & Hopper 2001, Pierrehumbert 2001, 2002, 2006, Seyfarth 2014, Sós-kuthy & Hay 2017). For example, Bybee and Scheibman (1999) show that in *don’t* contraction, the reduction process is most advanced with the most frequent context words, and that 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 and Scheibman (1999) collected and transcribed tokens of *don’t* from about three hours and forty-five minutes of ‘naturally occurring conversations’. In Table 1, which gives excerpts from Bybee & Scheibman 1999:581–82, 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.

According to Bybee and 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* con-

⁴ The BC consists of one-hour interviews with each of forty 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: twenty older (defined as age forty or more), twenty younger; twenty male, twenty female. The words and phones are aligned with sound waves, orthographically transcribed, and provided with broad phonetic labeling.

The CC is a subcorpus of the Origins of New Zealand English project. It consists of recorded and orthographically transcribed interviews. Speakers were 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 CC consisted of 1,087,113 words.

	[dõt, dõ]	[rõt, rõ]	[rã]	[ã]	TOTAL
PRECEDING					
<i>I</i>	16	22	38	12	88
<i>you</i>	7	7			14
<i>we</i>	2	6			8
<i>they</i>	1	3			4
lexical NP	5				5
...					
FOLLOWING					
<i>know</i>	2	8	24	5	39
<i>think</i>	7	6	6	1	20
<i>have</i>	1	7	1		9
<i>have to</i>	1	2	1		4
<i>want</i>	1	1	3		5
<i>see</i>	3	1			4
...					

TABLE 1. *Don't* variants by type of preceding and following item in data from Bybee & Scheibman 1999:581–82. Preceding and following contexts decrease in frequency from top to bottom; phonetic reduction increases from left to right.

trast 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.⁵

	FULL VOWEL	SCHWA
lexical sense	7	12
pragmatic function	1	17

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 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 (i) the fact that there are grammatical differences between the contracted and full forms: for example, *there's three men outside* vs. **there is three men outside* (see Dixon 1977, Nathan 1981, Sparks 1984, Kaisse 1985, Yaguchi 2010); (ii) that phonological rules that delete the onsets and schwas of specific auxiliaries cannot be assimilated to postlexical ‘rapid-speech phenomena such as deletion of flaps, coalescence of vowels etc.’ (Kaisse 1983:95); (iii) that the phonology of specific contractions cannot be assimilated to function-word reduction in general (Kaisse 1985); and (iv) that speech rate is not predictive

⁵ Applying a one-sided Fisher exact test to Table 2 to ascertain whether the odds ratio of vowel reduction cooccurring with the pragmatic function is reliably greater than 1, as predicted, yields p -value = 0.02545.

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 1969:730–31).

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.⁶

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, 2019, Bresnan & Spencer 2012, J. Spencer 2014, Barth & Kapatsinski 2017). 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 probability of *word*₁ appearing before *word*₂ (1a) in some language can be estimated from a particular corpus by the calculation shown 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.⁷

- (1) a. Conditional probability: $P(\textit{word}_1|\textit{word}_2)$
 b. Estimated: $\frac{\textit{count}(\textit{word}_1\textit{word}_2)}{\textit{count}(\textit{word}_2)}$
 c. Informativeness: $\log \frac{1}{P(\textit{word}_1|\textit{word}_2)} = -\log P(\textit{word}_1|\textit{word}_2)$

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 *word*₁ in the context of

⁶ In a corpus study of contractions in the Switchboard corpus (Godfrey & Holliman 1997), MacKenzie (2012:130, 149–55) 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). Research discussed below supports the effects of conditional probabilities of contraction with a host in the contexts of specific auxiliaries.

⁷ In information theory this quantity is known as SURPRISAL (Shannon 1948), which others have termed ‘local contextual predictability’ (e.g. Seyfarth 2014). Both terms, associated with the listener’s perspective, are avoided here with the more neutral term ‘informativeness’.

The weighted average of the surprisal of *word*_{*i*} over all contexts *j*, $-\log_2 P(\textit{word}_i|\textit{context}_j)$, is the INFORMATION CONTENT of the word, also termed ‘informativity’ and ‘average contextual predictability’ (Seyfarth 2014). When there is a fixed context of linguistic interest (such as an auxiliary), averaging over it adds nothing. Further, with a fixed context, the conditional probability of a host is proportional to its bigram probability with the auxiliary, studied by Krug (1998) and Bybee (2001, 2010).

the speaker's planned next word, $word_2$: the ratio measures $word_1$'s share of all tokens that precede $word_2$; it thus corresponds to $word_1$'s relative availability or activation in that context. The probability of the auxiliary given the PRECEDING word, $P(word_2|word_1)$, would presumably be more helpful to the listener, who does not have access to the speaker's planned next word.⁸ 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).

Figure 1 plots the relation between informativeness and contraction of present-tense *have* and *be* with pronominal hosts from the BC.⁹ 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 a steady decrease for present-tense forms of both *be* and *have*.

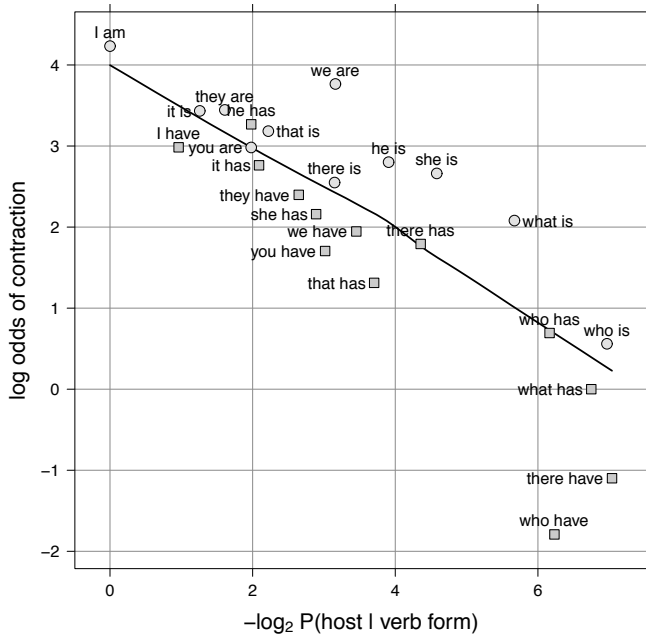


FIGURE 1. Relation between the log odds of contraction and the negative log conditional probability ('informativeness') of pronoun hosts in the context of verb forms in the Buckeye Corpus.

The *have* and *be* data sets are plotted with square and round point shapes, respectively, with a LOESS smoother showing the trend in the combined data.

⁸ 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 listener (e.g. Zhan & Levy 2018, 2019).

⁹ The data points in Fig. 1 represent 7,614 present-tense *be* forms and 805 present-tense *have* forms collected, respectively, by J. Spencer (2014) and the author from the BC by the orthographic transcriptions *have/ve*, *has/s*, *am/m*, *are/re*, and *is/s*. Instances in which the grammatical context did not permit contraction were excluded, following MacKenzie 2012. The remaining instances were checked against their phonetic transcriptions to ensure that orthographically contracted auxiliaries corresponded to phonetically asyllabic forms. Informativeness was calculated as in 1.

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 in 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 and clustered by similarity. 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.

In this model, speech perception involves the labeling of new instances based on their similarity to existing instances stored in memory, and speech production involves randomly selecting a target exemplar from the same space of stored memory instances; the production of that target is then added to the store of exemplars. This is the PERCEPTION-PRODUCTION LOOP, which dynamically affects language change by amplifying slight biases or changes over many iterations. For example, a slight but constant production bias toward lenition in each utterance can result in gradual sound changes in which more frequent words show a higher rate of change than less frequent (Phillips 1984, Bybee 2000): more frequently uttered words refresh their stores of lenited exemplars, while the exemplars of less frequently uttered words are selected less often as targets of production because of the greater impact of memory decay (Pierrehumbert 2001).¹⁰ Applied to multiword sequences, the model can account for the relation between usage and phonetic reduction (§1). Applied to host + tensed auxiliary sequences, the model can also account for the relation between usage probabilities and contraction (§2).

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

The mental lexicon stores both words and multiword fragments (Bybee & Scheibman 1999, Bybee 2010). Among the multiword fragments would be *you are*, the uncontracted sequence of function words that is functionally equivalent to *you’re* in grammatical structure (although they may of course differ in other properties, such as prosody, discourse context, and speaker style). Instances of both would have a common label at some level of grammatical labeling. In this way the mental lexicon would implicitly encode bigram probabilities and informativeness as activation levels of the various words and multiword fragments that are stored there.

¹⁰ 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 dynamic exemplar models further to incorporate the perceptual biases of the listener (Todd 2019, Todd et al. 2019).

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 that ‘word prevalence’—how many different people use a word—may be a better estimate of frequency effects on lexical decision times (Brysaert et al. 2016).

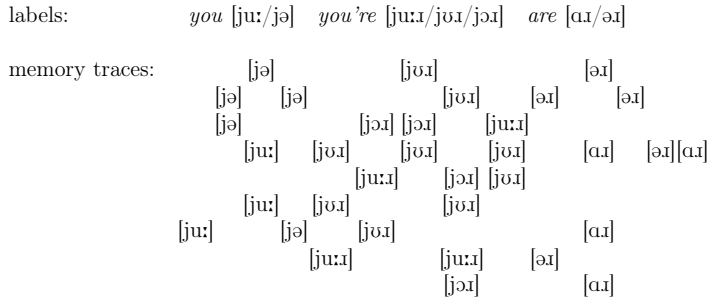


FIGURE 2. Visualization of tensed auxiliary contractions in an exemplar-dynamic model of the mental lexicon (Pierrehumbert 2001, 2002, 2006), which includes memory traces of speech events of varying activation levels (a function of the density of exemplars under each label as well as their recency).

In contraction, the short (asyllabic) allomorphs of the auxiliaries are phonologically incorporated into the final syllable of the host.¹¹ Assuming a production bias favoring the short allomorph parallel to the production bias favoring lenition, the crucial connection between high-probability (low-informativeness) host-auxiliary bigrams and higher incidences of contraction in speech production is then straightforward: under memory decay the clusters of more frequently uttered bigrams refresh their stores of contracted exemplars more often, while less frequently uttered bigrams are more temporally remote, lower in activation, and less likely to be selected as targets of production.

If highly probable contractions are lexically stored with phonetic detail, they should accumulate allophonic reductions as part of their long-term representations (§1). 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 online processes, such as that evident in other tokens of *don't*, as in *I don't inhale*, indicating that such reduction has accumulated in representation.

There is evidence that fits this expectation.

Wescoat (2005:471) gives various examples of ‘morphophonological idiosyncrasies’ among tensed auxiliary contractions, shown in Table 3. One of them is that ‘*I* [aɪ] may be pronounced [a], but only in association with *'ll* (*will*), yielding [aɪ]; moreover *you* may become [jə], but only when followed by *'re* (*are*), resulting in *you're* [jɔɪ]’.¹² Thus the reduced pronunciations are specific to individual pronoun-auxiliary sequences. He emphasizes that these pronunciations are not fast-speech phenomena: *I'll* [aɪ] and *you're* [jɔɪ] ‘may be heavily stressed and elongated’. In other words, their pronunciations are not merely online contextual adjustments to the phonology of rapid connected speech.

I'll	[aɪ/aɪ]	I'm	[aɪm/*am]	I've	[aɪv/*av]
you'll	[ju:l/*jɔɪ]	you're	[ju:ɪ/jɔɪ]	you've	[ju:v/*jɔv]

TABLE 3. Contrasting contraction-specific pronunciations from Wescoat (2005).

Diachronically, these pronunciations could plausibly derive from such online contextual adjustments of frequently repeated sequences (for example, the velarization or dark-

¹¹ Bybee (2002:124–25) demonstrates that spoken usage frequencies favor encliticization over procliticization of the asyllabic auxiliary.

¹² 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.

ening of /l/ in *will* and the laxing of immediately preceding unstressed vowels, yielding *we'll* ['wi:əl, wɪl]). 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 online 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 raw frequency), across lexicons from a variety of languages:

One likely mechanism for how the lexicon comes to reflect predictability is that information content is known to influence the amount of time speakers take to pronounce a word: words and phones are given shorter pronunciations in contexts in which they are highly predictable or convey less information [references omitted]. If these production patterns are lexicalized, word length will come to depend on average informativeness. (Piantadosi et al. 2011:3528)

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).¹³

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,¹⁴ and without adequate representations of the grammatical context they cannot account for constraints on contraction (§7).

The following three subsections summarize those findings most relevant to the present study. For these it is useful to distinguish between unstressed syllabic and asyllabic forms of the tensed auxiliaries as in Table 4, adapted from Inkelas & Zec 1993 and Wescoat 2005, who follows Sweet 1890:14–16.¹⁵

4.1. 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.

¹³ Seyfarth (2014) discusses this and possible alternative models of the effects of 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. 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 accumulation of lexically specific phonetic, semantic, and pragmatic features of the kind found by Bybee and Scheibman (1999) (see Tables 1 and 2).

See Bybee & McClelland 2005 for discussion of a distributed connectionist alternative model and Ambridge 2020 for a broad discussion of exemplar theories and alternative models in an acquisition context.

¹⁴ Barth and Kapatsinski do, however, analyze contraction by broad construction type such as copula, future, and progressive (Barth 2011, 2019, Barth & Kapatsinski 2017).

¹⁵ In the present study [ə] represents the stressless mid central vowel, and [i] represents a slightly higher unstressed vowel.

	FULL (‘STRONG’)	UNSTRESSED SYLLABIC (‘WEAK’)	ASYLLABIC (‘ENCLITIC’)
are	[ɑɪ]	[əɪ/ɪ]	[ɪ]
am	[æm]	[əm]	[m]
had	[hæd]	[(h)əd]	[d]
have	[hæv]	[(h)əv]	[v]
has	[hæz]	[(h)əz]	[z/s]
is	[ɪz]	[ɪz/əz]	[z/s]
will	[wɪl]	[əl/ɪ]	[l]
would	[wʊd]	[(w)əd]	[d]

TABLE 4. Strong, weak, and enclitic forms of the tensed auxiliaries.

The right context of both syllabic and asyllabic reduced auxiliaries requires that the auxiliary be followed by a stressed word, as 2a,b illustrate.¹⁶

- (2) a. They are/*’re __ . [ðeɪ ɑɪ/*’ðeɪ.əɪ/*’ðeɪ.ɪ]
 b. They are/’re here. [ðeɪ ɑɪ/’ðeɪ.əɪ/’ðeɪ.ɪ]

The stressed word need not be adjacent to the auxiliary. In line with Labov’s (1969) observations as well as the corpus evidence of MacKenzie (2012:79–82), *is* reduces and contracts before the nonadjacent stressed verb *doing* in 3a, but not before unstressed *it* alone.¹⁷

- (3) a. That bird, what’s it doing __ ? [ɪwʌts.ɪt ’duɪŋ]/[ɪwʌt.əz/ɪz.ɪt ’duɪŋ]
 b. *That bird, what’s it __ ? *[ɪwʌts.ɪt]/*[ɪwʌt.əz/ɪz.ɪt]
 (cf. ... , what is it?/what’s it?)

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

- (4) I don’t know where the party is [ɪz/*iz/*z] tonight.

Since the publication of a squib that excited decades of work on the syntax of auxiliary contraction (King 1970), many linguists have continued to judge contraction to be blocked before prefocus gaps and ellipsis medially within the complement phrase, as in the ‘comparative subdeletion’ (5) and ‘pseudogapping’ (6) examples from Inkelas & Zec 1993. In these examples, the words in small caps are uttered with pitch accents, indicating parallel foci contrasting both subject and object.

- (5) KAREN is a better DETECTIVE than KEN is/*’s an ARCHEOLOGIST.
 (6) JOHN’S playing ROULETTE, and MARY is/*’s BLACKJACK.

To account for the apparent ungrammaticality of contraction in such examples, many analyses have hypothesized that contraction is blocked within the verb phrase before a medial syntactic gap or ellipsis for various reasons (e.g. Bresnan 1971, Kaisse 1983,

¹⁶ Following Wescoat 2005, the dot ‘.’ marks a syllable boundary.

¹⁷ Contractions without a stressed complement occur where an utterance is interrupted, incomplete, or disfluent. An authentic CC example shows *it’s* before a pause (‘.’) and restart: *it’s . it’s not the be all and end all*. Contraction before final unstressed *it* also occurs in some fixed expressions. An authentic example from CC is *I’d be surprised if Peter and who’s your what’s it didn’t get in*, referring to someone or something whose name has been temporarily forgotten (also spelled *whosie-whatsit*). A similar example is *howsit/howzit*, a New Zealand slang greeting. Another likely fixed expression is *what’s it* referring to something that the speaker is trying to remember, as in this authentic example from CC: *this um – what’s it . it was a TR250*.

1985, Inkelas & Zec 1993, Anderson 2008).¹⁸ But Selkirk (1984) suggests that such phrase-medial sites of contraction actually allow variable contraction in usage, citing 7.

- (7) Looks as good as it's ___ fun to play. (Selkirk 1984:443, n. 25)

And there are authentic examples on the Web in support of this suggestion.

- (8) a. 'But I know **he's a better runner than he's a biker**,' Lope 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!

Interestingly, these examples differ from those judged ungrammatical (5 and 6) in that instead of contrasting two pairs of foci—the subject and object complement of each clause—they contrast only one: the complement; the subject of each second clause is anaphoric and not in contrast. Yet the examples do not differ in the relevant prefocus syntactic structure, so the hypothetical gap cannot be what prevents contraction.

To account 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 prefocus is one of a suite of stylistic metrical options that speakers may use to highlight prosodic and structural parallelism in constructions like those in 5 and 6 above. That is the position adopted in the present study.

Inkelas and Zec (1993) also point to examples like 9a,b, where contraction can occur directly before a gap (a), provided that a stressed complement word follows (contrast b).

- (9) a. I don't know how much there is/'s ___ left in the tank.
 b. *I don't know how much there is/*'s ___ .

Similar examples occur on the Web, showing contractions adjacent to the extraction sites.

- (10) 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 on-line?
 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**.

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 a possessive enclitic in 11b, a weak object enclitic in 11c, and an untensed auxiliary enclitic in 11d.

- (11) a. Who's very polite? *Tom's. (= Tom is)
 b. Whose child is very polite? Tom's. (= Tom's child)
 c. Kill 'em. ['kɪl.ɪm]
 d. I might've. ['maɪt.əv]

¹⁸ The gap in 5 is supposed to correspond to an implicit degree modifier such as (*how good*) or (*that good*) an *archeologist* (Bresnan 1973).

4.2. ENCLISIS WITH THE LEFT CONTEXT. While sharing their metrical dependence on a stressed complement in the right context, the asyllabic and unstressed syllabic auxiliaries diverge with respect to the left context. Specifically, the asyllabic tensed auxiliaries form a PHONOLOGICAL WORD with their hosts to the left, unlike their syllabic counterparts. The phonological wordhood of tensed auxiliary contractions is supported by (i) the progressive voicing assimilation of *ʒ* with the final segment of the host, together with (ii) the absence of pausing and interruptions between the host and the contracted auxiliary.

Examples 12 and 13 illustrate the phenomenon of voicing assimilation (i). The choice of the specific pronunciation of *ʒ* depends on the phonology of the host. Morphophonologically, *ʒ* 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.

- (12) a. PLURALS: peats [s], reds [z], losses [əz]
 b. PRESENT TENSE: bleats [s], shreds [z], tosses [əz]
 c. *ʒ* CONTRACTIONS: Pete's [s] here, Fred's [z] here, Ross's [iz] here

The contrast with arbitrary adjacent syntactic words shows that the voicing assimilation and epenthesis are word-internal effects specific to contractions with the auxiliary *ʒ*.

- (13) a. Pete sang [s] and Fred sang [s/*z].
 b. Fred zigged [z] and Pete zagged [z/*s].
 c. Ross zig-zagged [z/*əz].

As for (ii) above, the authentic examples in 14a–c are provided by MacKenzie (2012:76–79) to illustrate that contraction of the auxiliary is not found in ‘pseudo-clefts’, ‘*th*-clefts’, and ‘*all*-clefts’.

- (14) a. What I'm talking about is [ɪz] the people over here and over here and across the street.
 b. Well, the problem is [ɪz], that most of the record players now will not play them.
 c. All I know is [ɪz] I didn't vote for him.

Examples 14a–c can be thought of as SPECIFICATION CONSTRUCTIONS in that the postauxiliary 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 and Zec (1993:243, 245) propose a phonological explanation that could apply to such specification constructions, as well as other authentic preauxiliary contexts that block extraction from MacKenzie 2012¹⁹ and constructed examples from Kaisse 1979, 1983, 1985. Inkelas and Zec (1993) assume that English auxiliary clitics form a phonological word ω with a phonological word to their left. Then they assume with Sells (1993) that certain focused syntactic constituents are set off by a phonological or intonation phrase boundary, which prevents auxiliary enclitization.²⁰ How this proposal would apply to 14 is illustrated in 15.

- (15) *{What I'm talking (**about**) {*s*} _{ω} the people over here ... }

If contractions are enclitics on their hosts to the left, they cannot be interrupted by pauses or by the prosodic boundaries of certain focused or dislocated syntactic phrases.

¹⁹ These would include parentheticals, adverbs, and preposed prepositional phrases (locative inversions).

²⁰ They assume that the strong prosodic boundary is obligatory, but for other speakers it appears to be an optional variant. Individual or stylistic variability in the strength of prosodic boundaries would explain contrasting grammaticality judgments of constructions like *Speaking tonight is! *ʒ our star reporter* (cf. Inkelas & Zec 1993:245 and Anderson 2005:71).

In contrast, the tensed weak SYLLABIC auxiliaries are not enclitics but are phonologically dependent on their RIGHTWARD phrasal context only (Selkirk 1984, 1996, Inkelas 1991, Inkelas & Zec 1993). Hence pauses and strong prosodic boundaries can separate them from the preceding word.

(16) {What I'm talking about} {is [iz] the people over here and over here and across the street}

(17) {They—bicycle cranks, I mean—are [əɪ] 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, while both the syllabic and asyllabic forms are 'proclitic' in a purely metrical sense (cf. Bresnan 1971, Wilder 1997), only the asyllabic form also encliticizes to its preceding host.

4.3. RESTRICTIVE AND NONRESTRICTIVE AUXILIARIES. While all asyllabic tensed 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 2005.²¹

(18) Classification of asyllabic forms

	restrictive	nonrestrictive
are	're	
am	'm	
had		'd
have	've	
has		's
is		's
will	'll	
would		'd

According to Wescoat (2005), the restrictive asyllabic auxiliaries contract only with pronoun and WH-proform hosts, while other asyllabic auxiliaries are not restricted in this way. His examples 19a–e show restrictive asyllabic auxiliaries with pronoun and WH-proform hosts.

- (19) a. I'll help. [aɪl]
 b. We're a big group. [wi:ɪ]
 c. They've gone. [ðeɪv]
 d. I'm happy. [aɪm]
 e. How've you been? [haʊv]

Wescoat (2005) constructs minimal pairs to 19 using monosyllabic nonpronoun hosts, which he judges ungrammatical when pronounced with asyllabic contraction.²² Exam-

²¹ Wescoat follows A. Spencer's (1991:383) classification of 'd as restrictive, but notes that it contracts with nonpronoun 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.

²² Wescoat's theory is compatible with a range of varying judgments, because it depends on lexical features of the host. For example, if Wescoat (2005) had categorized *so* as a proform rather than an adverb, it could allow contraction with restrictive asyllabic auxiliaries, and indeed a referee made this judgment. Further variations are discussed in §9.

ples include *Ai'll help* [aɪ.lɪ]/*[aɪl], *The Cree're a big group* [kɹi:ɹɪ]/*[kɹi:ɹɪ], *The Au've been polled* [aʊ.əv]/*[aʊv], and *So am I* [sɒ.əm]/*[sɒəm].²³

The nonrestrictive asyllabic auxiliaries corresponding to *is*, *has*, *had*, and *would* can all contract with both pronoun and nonpronoun hosts in some varieties of American English, as the following examples slightly adapted from Wescoat 2005 illustrate.

- (20) a. It's gone/going. [ɪts]
 b. Pat's gone/going. [pæts]
 (21) a. She'd seen it. [ʃi:d]
 b. Lee'd seen it. [li:d]
 (22) a. I'd have seen it. [aɪd]
 b. Bligh'd have seen it. [blaɪd]

There is a further syntactic difference between restrictive and nonrestrictive auxiliaries, illustrated by examples 23a–c from Wescoat 2005: the hosts of the former cannot be conjuncts or occur embedded within a larger subject phrase.

- (23) a. [She and **I**]'ll help. [aɪ.lɪ/*aɪl/*aɪ]
 b. [The people beside **you**]'re going. [ju:ɪ/*ju:ɪ/*jʊɪ/*jʊɪ]
 c. [The people who helped **you**]'re kind. [ju:ɪ/*ju:ɪ/*jʊɪ/*jʊɪ]

In contrast, the following authentic spoken examples—24a,b from the CC and 24c,d from the BC—illustrate that nonrestrictive *'s* can contract with noun hosts that are dependents of the subject of the auxiliary and conjuncts.

- (24) 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

Although authentic examples are rarer, other nonrestrictive auxiliaries may not contract as freely as *'s*. Example 25 shows the single instance of *'d* contracted with a nonpronoun host in 2,890 occurrences of contracted and uncontracted *did*, *had*, and *would* in the BC.²⁴

- (25) *Wexner Center 'd* [ˈsɛntə-d] <SIL> be one of my primary ones.

Judgments of constructed data are uncertain, but 26a–b suggest that both *would* and *had* can contract with a host embedded within a larger subject phrase, at least in the author's speech.

- (26) a. [Everybody in my **family**]'d agree. [ˈfæm(ə)lɪd] d < would
 b. [Everybody in my **family**]'d agreed to it. [ˈfæm(ə)lɪd] d < had

In sum, beyond their shared prosodic and metrical properties, the contracting auxiliaries appear to differ in their selectivity for the host words and their restrictiveness toward host phrases. The restrictive auxiliaries require that the host be a subject pronoun or WH-proform not embedded within a larger subject phrase. The nonrestrictive 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

²³ According 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).

²⁴ Here <SIL> labels a silent region during a speech disfluency of some kind, such as a restart or hesitation (Kiesling et al. 2006:19).

enclisis, as Wescoat (2005) argues. When viewed as purely phonological phrasings of two adjacent nonconstituent 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 and Sag’s (2001) HEAD-DRIVEN PHRASE STRUCTURE GRAMMAR (HPSG) analysis, drawing on A. 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, as the examples in 27 illustrate.

- (27) a. ***[They’re and you’re]** going.
- b. **I’m** looking forward to seeing you] and [will be there on Sunday].
- c. **You’ll** do what I say] or [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.²⁵

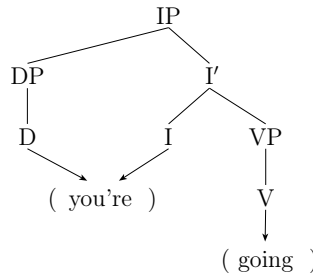


FIGURE 3. Relation between c-structure and tensed auxiliary contraction under lexical sharing (Wescoat 2005).

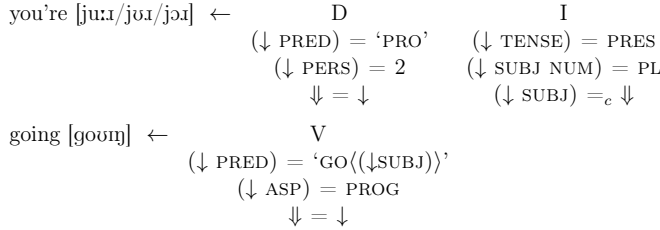
In Fig. 3 the arrows pointing to words represent a formal mapping from syntactic constituent structures (c-structures in LFG) to the lexical items that instantiate them—their LEXICAL EXPONENTS, in Wescoat’s (2005) terms. As usual in LFG, the c-structure represents the ‘surface’ syntactic groupings of words, while the ‘deeper’ relations and dependencies are provided in a parallel functional structure (f-structure) that bears many similarities to dependency grammar graphs (Mel’čuk 1988, Bresnan 2016). The surface words themselves provide most of the global functional information in the form of relational features that give rise to descriptions of the f-structure context of the word. Language-particular c-structure configurations provide what structural information

²⁵ The particular category names are not important; here Wescoat follows the c-structure theory outlined in Bresnan 2001 (see 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.

about linguistic functions there may be in a given language, which in the case of configurational languages like English is fairly redundant (Bresnan et al. 2015).

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, emphasis in original). In the lexicon these restrictive contractions are associated with adjacent syntactic terminal categories and may specify item-specific phonology and functional restrictions, as illustrated in 28.²⁶ In 28 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.

(28) Lexical entries for the structure in Fig. 3



Figures 4 and 5 provide extensional visualizations of the structures and relations specified by these lexical entries. The visualization in Fig. 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 28. These fragmentary lexical structures are merged and integrated in specific syntactic contexts, such as that in Fig. 3.

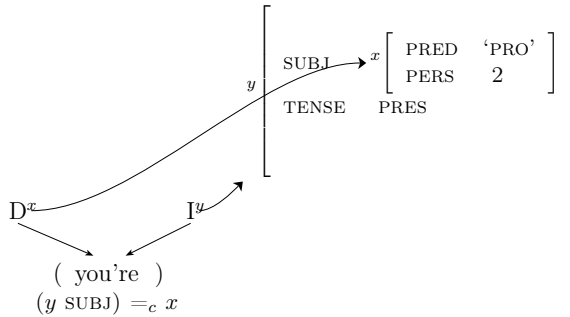


FIGURE 4. Information specified by the shared lexical entry in 28: the curved arrows represent mappings from c-structure terminals to f-structures, and the straight arrows are mappings from the c-structure terminals to their shared lexical exponents.

²⁶ The ‘down’ arrows in 28 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 \Downarrow 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 28, 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 $\Downarrow = \downarrow$ identifies the functional structure of the host *you* 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

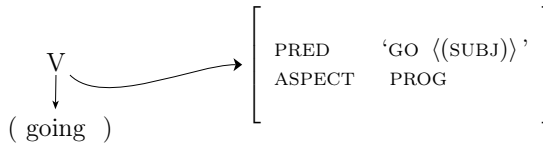


FIGURE 5. Information specified by the verb lexical entry in 28: again, the curved arrow indicates the mapping from the c-structure terminal to its f-structure, and the straight arrow maps the c-structure terminal to its lexical exponent.

Wescoat shows that the correct f-structure for Fig. 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 cohead, lexically sharing the contraction *you're*, which provides their substantive features. (See Wescoat 2005 for more details.)

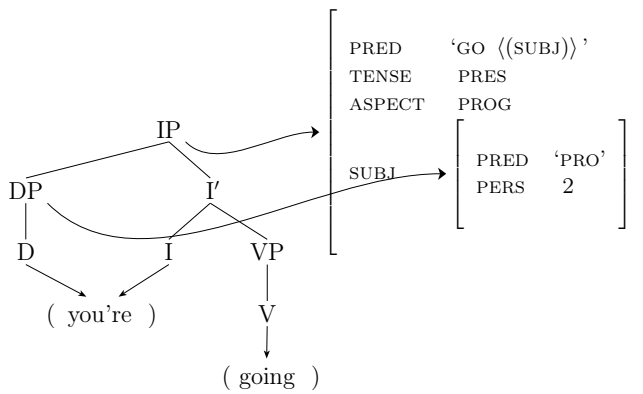


FIGURE 6. C-structure to f-structure links for the structure in Fig. 3 given the lexical entries in 28.

The main prosodic, syntactic, and morphophonological properties shared by all tensed auxiliary contractions follow from this analysis. (i) Host + auxiliary contractions cannot be conjoined to each other as in 27a because they are not c-structure units. (ii) The coordination of two auxiliaries together with their verb phrases, despite the first being contracted with the subject as in 27b,c, is simply I' coordination, as Wescoat points out. (iii) 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. (iv) 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 (see, for example, Selkirk 1996).²⁷

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.

²⁷ A referee points out that contrary to this assumption, Levelt et al.'s (1999) lexical access model is designed to allow the phonological word to cross lexical word boundaries. Their evidence comes from resyllabification between verbs and their unstressed pronoun objects: *escort us* syllabified as *es.kor.tus*, and *understand it* as *un.der.stan.dit* (Levelt et al. 1999:20, 31). However, there is much evidence that these unstressed pronominal

As for the syntactic properties that distinguish restrictive from nonrestrictive auxiliaries (§4), those are captured as restrictions on the common f-structure of the lexically shared host + auxiliary. Examples occur in the lexical entries making use of the metavariable \Downarrow (n. 26). In 28, 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 SUBJ function of the atomic auxiliary I f-structure. This constraint immediately accounts for syntactic restrictions illustrated in 23a–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 its inverted position before the subject (denoted C) as the extended cohead of its clause (Bresnan 2001:103, Bresnan et al. 2015).

(29) Lexical entry for an inverted auxiliary contraction

how've [haʊv]	←	ADV	C
		(↓ PRED) = 'HOW'	(↓ TENSE) = PRES
		↓ = ↓	(↓ ASPECT) = PERF
			(↓ FOCUS) = _c ↓

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

Thus the theory of lexically shared clitics adopted here improves on preceding purely prosodic and purely morphological theories of restrictive auxiliary contraction by analyzing them as lexical units whose adjacent components simultaneously retain some syntactic independence in c-structure.

6. LEXICAL SHARING OF NONRESTRICTIVE CONTRACTIONS. Wescoat (2005:482) proposes extending the theory of lexical sharing from restrictive 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 \int [z/s/əz] (*is* or *has*) to a host, yielding a lexical-sharing structure; the host may be anything, the attachment of \int [z/s/əz] 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 \int contractions. Example 30 shows a lexical schema for contracted *is*.²⁸ It differs from the entry for *you're* shown in 28 in that here the restriction (↓ SUBJ) =_c ↓ is absent and the host and its category are unspecified. This schema can be viewed as Wescoat's (2005) 'lexical process' for attaching nonrestrictive \int to hosts.

(30) Lexical schema for contracted *is*

x's [... z/s/iz]	←	X	I
		...	(↓ TENSE) = PRES
		↓ = ↓	(↓ SUBJ NUM) = SG
			(↓ SUBJ PERS) = 3

objects in English are not independent lexical words, but enclitics (see, for example, Abercrombie 1961, Selkirk 1972, 1996, Zwicky 1977), so they would not be true examples of resyllabification across lexical word boundaries. Selkirk (1996) analyzes them as 'affixal prosodic clitics'. Note that while all lexical words are phonological words, some phonological words might be produced from syntactic enclisis (§7).

²⁸ The generalization to contractions of inverted \int would allow C as an extended head as well as I; see discussion of 29.

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 31. Note that the lexical entry has the schematic structure in 30, which requires adjacency in c-structure between the host and auxiliary categories. As with other instances of lexical sharing, the host and contracted auxiliary that satisfy the lexical schema form a phonological word.

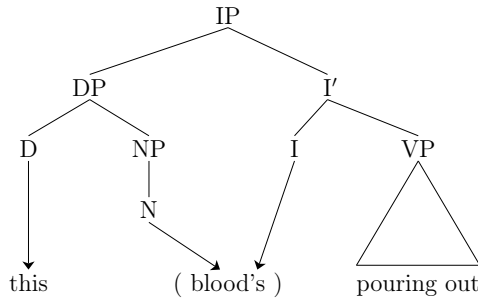


FIGURE 7. An example c-structure with *'s* contraction under lexical sharing.

(31) Lexical entry for the contraction *blood's* in Fig. 7, derived from schema 30

blood's [blʌdz]	←	N	I
		(↓ PRED) = 'BLOOD'	(↓ TENSE) = PRES
		↓ = ↓	(↓ SUBJ NUM) = SG
			(↓ SUBJ PERS) = 3

Figure 8 shows how the structure in Fig. 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 coheads, just as I and VP are coheads. Because the f-structures of coheads unify, the features of the NP dominating the host N are unified with the features of the proximate demonstrative D *this*.

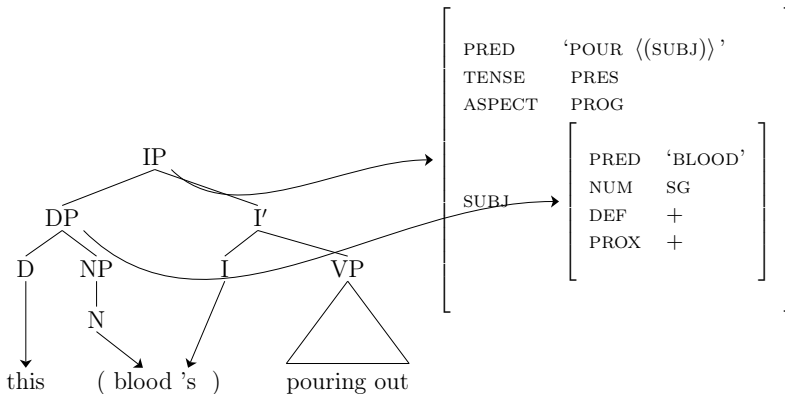


FIGURE 8. C-structure to f-structure links for the structure in Fig. 7 given the lexical entry in 31 and general LFG principles of structure-function mapping (Wescoat 2005).

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 32a–c are authentic examples from the Web, selected

7. A HYBRID MODEL. Combining the formal grammar and the usage-based mental lexicon reviewed in previous sections into a hybrid model is the subject of the present section. As mentioned at the outset, the aim is not to present a detailed formalization, but to describe at a high level how the architecture of the dynamic exemplar model discussed in §3 could combine with the formal grammar of the present study to explain the main empirical findings of both usage-based and formal lines of research on tensed auxiliary contraction.

In the present framework a hybrid model of SYNTACTIC production (excluding higher-level discourse context and semantics) would use f-structures as input representations, lexical entries as labels of memory-trace clouds, and the ordered lexical exponents of c-structures as outputs. These concepts are illustrated in Figures 10 and 11 for the production of *you're* or *you are*.

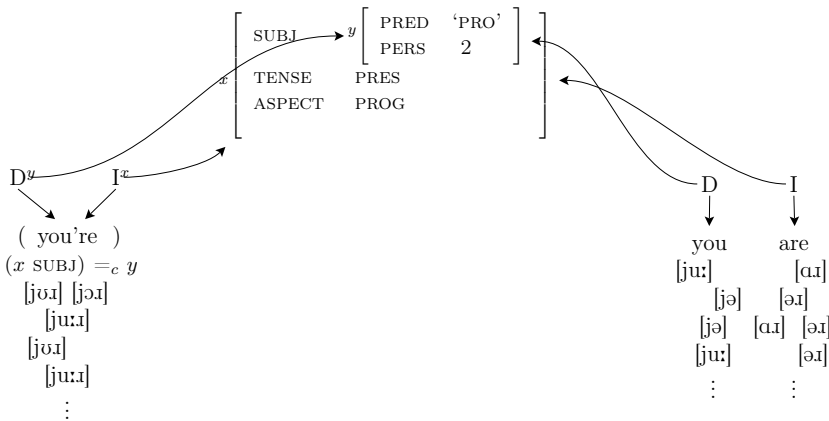


FIGURE 10. A visualization of a production input as an abstract phrase or sentence plan (an f-structure) linked to words in the mental lexicon whose functional schemata match it. The words label clouds of memory traces, from which an exemplar is randomly selected as the target of production.

Figure 10 illustrates an input to speech production at the syntactic level as an abstract plan for a phrase or sentence. The plan is represented by a functional structure for a second-person pronoun subject of a clause in the present progressive. Activation of this f-structure would activate the words that are linked to it in the mental lexicon: *you're*, *you*, and *are*. These are the labels most similar to the input in their relational features—specifically, the words whose functional schemata in their lexical entries can be instantiated to match the input f-structure. (Compare the extensional visualizations of the functional schemata of lexical entries in Figs. 4 and 5.)

These lexical entries would each label a cloud of memory traces, like the illustration in Fig. 2, which uses orthographic words as labels. The word clouds of *you* and *are* would be bound together by their links to the same input f-structure and as a set would serve as a composite label for the union of the word clouds for *you* and *are*. Thus the hybrid model incorporates both contractions and their uncontracted multiword equivalents in the mental lexicon (cf. §3).

An exemplar would be randomly selected as a target of production from the union of the clouds of *you're* and the composite label *you are*. If nothing differentiates them in the input context, the contracted and uncontracted variant exemplars would both be possible selections as targets of production (with their likelihood of selection weighted by the relative activation of their clouds).

To produce a syntactic output from the selected production target, the syntactic production process would fit the winning exemplar into the phrase patterns of English in accordance with its lexical entry or entries so that it corresponds to the input f-structure. Details of generation and parsing are outside the scope of the present study,²⁹ but the syntactic output of the example input could be one of the alternative strings of ordered lexical exponents in the c-structures shown in Figure 11.³⁰

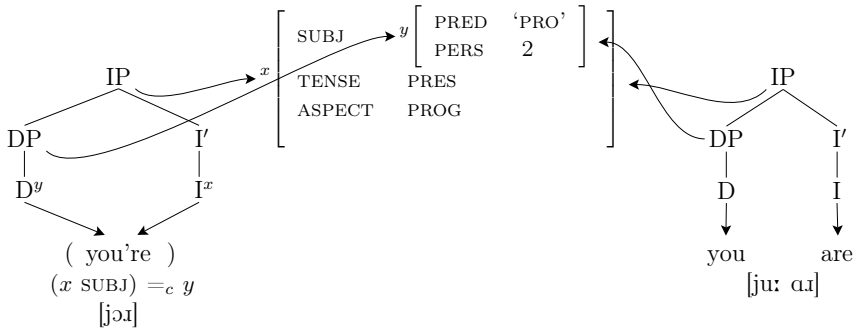


FIGURE 11. A visualization of alternative production outputs.

Figures 10 and 11 represent a synchronic model of production, but the diachronic applications of the dynamic exemplar model elsewhere (§3) lead to the question of why the syntactic structure on the left-hand side of Fig. 11 arises as a variant of that on the right.

Observe that the contraction cannot simply be a phonetically fused sequence of adjacent words or allomorphs, as described in §3, because the fusion does not occur everywhere that the sequence of the adjacent pronoun and auxiliary occurs. Recall 23b–c, for example: [*The people beside you*]'re going, pronounced [jʊ:ɹɪ] but not [*jʊɹɪ/*jɹɪ]. Thus what is lexically stored is not merely a sequence of words and allomorphs, but fragments of syntactic structures they occur in with their local relations and dependencies, as visualized in the left side of Fig. 10. These syntactic fragments can enter into conjuncts parallel to uncontracted phrases, as in *You*['re gonna do what I say] or [*will suffer the consequences*] (cf. 27b,c). And they share the rightward metrical dependence of unstressed *are* in uncontracted phrases (cf. 2–4).

Consequently, at the syntactic level the lexical storage of high-probability restrictive auxiliary sequences like *you* and *'re* as units must include the storage of the fragments of syntactic structure they occur in. This is what lexical sharing does: it specifies the contracted sequence *you're* as a sequence of word categories that share a common functional structure in which *you* is the required subject of *are*.

Observe that the syntactic restriction of the asyllabic auxiliary *'re* to subject pronoun hosts singles out the syntactic context that has the highest share of token frequencies of cooccurrence with the auxiliary.³¹ For this reason, lexical sharing as a formal construct

²⁹ Wedekind and Kaplan (2012) discuss various computational linguistic generation algorithms for LFG.

³⁰ The curved arrow mappings from IP and DP to the f-structure in Fig. 11 arise because in the syntax the f-structure of a node is identified with that of its head or cohead (Bresnan et al. 2015). Alternative theories of c-structure could of course be adopted, with varying degrees of flatness or hierarchical structure and finer- or coarser-grained part-of-speech categories (n. 25).

³¹ For evidence, note that in the context preceding an adjacent tensed auxiliary *is* or *'s*, subject pronouns far outnumber other host phrases (§8).

can be viewed as a grammaticalization of high-probability syntactic distributions in usage, just as the allomorphs of *you* and unstressed *are* can be seen as a grammaticalization of high-probability pronunciations.

How should the model handle *is* contractions? With restrictive auxiliaries like *are*, *'re* in Fig. 10, the high probability of cooccurrence with their pronoun hosts leads to repeated phonetic reductions of the host that become lexicalized over the long term, providing independent support for the lexical storage of the host + auxiliary combinations as units. But with the nonrestrictive auxiliary *'s*, evidence of such long-term phonetic reductions of hosts is lacking. At the same time, authentic examples like 24a–d and 32a–c suggest that this auxiliary lacks all syntactic constraints on its host except adjacency (but see §9). The schematic shared lexical entry for *'s* (30) expresses both of these properties: it neither selects a specific lexical host nor imposes the requirement that the host be its subject or have any relation other than being an adjacent word category to the left. Imported into the model of the mental lexicon, this entry would essentially provide a lexical label for the clitic *'s* without a specific host, simply as an allomorph of *is*.

In the mental lexicon, the clitic *'s* would label a cloud of memory traces just as the uncontracted *is* does (cf. *are* in Fig. 10). Then its activation, selection, and output production would proceed like that of *is*, except that its lexical entry would specify an adjacent host of any category to its left. The output production process would cliticize *'s* onto its host in accordance with its lexical entry (30), forming a phonological word (cf. Inkelas 1991, Inkelas & Zec 1993), and then fit it into the c-structure patterns of English that correspond to the input.

By itself, this analysis of contracted *'s* would yield a free and productive choice of *'s*, like *is*, for any adjacent host. Productions of *is* contraction could take place with novel hosts. And if that were the whole story, the probability of *is* contraction would be independent of the word serving as host. Instead of being conditioned on the joint occurrence of host with *is* forms, it would be roughly constant across hosts, dependent only on the proportions of the clitic *'s* and the syllabic forms of *is*.

However, if speakers produce the host word adjacent to the clitic *'s* sufficiently often, the sequence could become a lexically stored unit, parallel to *you're* in Fig. 10. The assumption needed for unit formation to occur is the perception-production loop: what is produced is also perceived and stored, and that will include generated productions. Given memory decay, infrequent and temporally remote stored combinations would become inaccessible as units and require generation by cliticization. In contrast, frequent and recent composite exemplars, like *he's* or *Mum's*,³² could become increasingly accessible established units. In this way, *is* contractions could in principle have dual sources either as stored units with specific hosts or as freely generated cliticizations,³³ and would show increasing contraction with sufficiently increasing frequencies of cooccurrence of host and auxiliary.

In sum, the hybrid model incorporates the usage-based explanation for the fact that the frequency of cooccurrence of host + auxiliary correlates with their likelihood of contraction (§3). But because the labels of its exemplar clouds are lexically shared representations of formal grammar that have well-defined mappings to syntactic input and output structures, it also entails the grammatical properties that restrictive auxiliaries share with

³² As the next section shows, *Mum* is one of the most frequent nouns that occurs before *is* or *'s* in the CC.

³³ Cf. Lowe's (2016) lexical-sharing analysis of genitive *'s*, and n. 52.

equivalent uncontracted phrases (§§4–5). Hence, the hybrid model has broader explanatory scope than either of its usage-based or formal-grammar-based components alone.

8. A CORPUS STUDY. This section presents corpus evidence for the hybrid model of tensed auxiliary contraction. For the corpus study, all instances of *is* and orthographic *ʼs*, totaling 20,095, were collected in 2015 from the CC transcriptions at the New Zealand Institute for Language, Brain, and Behaviour and manually checked and cleaned.³⁴

The formal theory of lexical sharing rules out contraction when an adverb separates the host from the auxiliary, because lexically shared elements must be adjacent.³⁵ Furthermore, given that lexically shared entries form prosodic words (§5), the grammar also rules out interruption of contractions by pauses and strong prosodic boundaries (see 15–16 and n. 20). Hence the formal component of the hybrid model of tensed auxiliary contractions entails that contractions should not occur in grammatical contexts with preauxiliary adverbs and in specification constructions. The CC data provide empirical support. In the twenty-six utterances with an adverb separating the host and auxiliary, there are no contractions. Examples are (i) *yep uh my girlfriend **actually is** a checkout supervisor*; (ii) *I think two families **at the moment is** just ...*. Likewise, in the 194 utterances consisting of specification constructions, there are also no contractions. Examples (prosodic boundaries added) are (i) *{the best thing to **do**} {is go ride your bike} and just keep doing it*; (ii) *cos {all you **do**} {is you just look down the machine hole}*. In contrast, the overall average rate of contraction in the data set containing only variable (optional) contractions ($n = 11,719$) is 91.3%, and the average of nonpronoun hosts ($n = 1,398$) is 56.5%. Thus it is highly unlikely that these theoretically expected failures of contraction in the left context are accidental.

As for the right context, the formal component of the hybrid model ties the contracted auxiliary to its own c-structure node, taking a complement phrase that can satisfy the rightward metrical dependence of auxiliary contraction (§§5 and 6). Hence, when the auxiliary is in positions without a following complement phrase, such as VP-final or clause-final positions, it should not contract. In the CC data, there are 189 instances where the auxiliary is final in the verb phrase or clause. Of these, only two are contracted. Two of the expected uncontracted examples are (i) *one o'clock I think **the exam start is** .*; (ii) *yeah **it is** definitely*. The two unexpected contracted instances are (i) *that's* and (ii) *yeah that's* (with no other context). Even with these two exceptions, the contrast with the variable portion of the data set is highly unlikely to be accidental.³⁶

In the dynamic exemplar lexicon of the hybrid model, contractions are theorized to be a function of the usage probabilities (measured as informativeness) of specific host-auxiliary sequences—even in the case of unrestricted *is* contraction, which formally re-

³⁴ With raw data collected by Jen Hay, research assistant Vicky Watson manually checked a sample against the audio files for transcription accuracy and marked data to eliminate transcription errors and instances of *ʼs* representing *has*, *does*, or possession. In addition, Watson followed MacKenzie 2012:65–90 in marking grammatical contexts that prevent or require contraction, including instances of *is/ʼs* followed by *not* or *nʔ*; hosts with final sibilants, which do not occur with the asyllabic auxiliary *ʼs*; clause-final and phrase-final occurrences of the auxiliary; specification constructions like 14a–c; and pauses or adverbs intervening between host and auxiliary.

³⁵ The reason is that the mapping from c-structure to lexical exponents is a homomorphism preserving precedence but not dominance (Wescoat 2002, 2005).

³⁶ The orthographic transcriptions do not include information about stress, so corpus evidence parallel to the judgment evidence in 3a,b is lacking.

sembles simple generative cliticization. Therefore it is of great interest to discover whether this theoretical prediction is borne out empirically.

To weigh the effects of informativeness in the corpus data, one must focus on the portion of data where informativeness could affect the probability of contraction—that is, on data where contractions are not already ruled out (or in) by the grammar itself. To this end, the rightward and leftward grammatical contexts just discussed were excluded from the final data set, along with the other invariable contexts from the linguistic literature that were empirically supported by MacKenzie (2012:65–90) in her corpus study. The resulting data set of variably contracting instances yielded 11,719 total observations from 412 speakers (mean instances per speaker = 28, standard deviation = 23) and 758 unique nonpronoun hosts.

Hosts in this data set were labeled semiautomatically as pronouns/nonpronouns, and their informativeness in the context of the auxiliary *is/s* was calculated from ngram statistics provided by Jen Hay and Robert Fromont for the entire CC of 1,087,113 words. Unsurprisingly, nonpronouns have higher informativeness than pronouns before the tensed auxiliary, and the hybrid model predicts that their likelihood of *is* contraction should be lower. The data bear out this expectation:

- Of 11,719 total observations of variable full and contracted *is*, 88% follow adjacent subject pronouns and 12% follow adjacent nonpronouns.
- Contraction appears with 96% of the former and 56.5% of the latter observations.

Second, among nonpronoun 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 CC. The nonpronoun hosts having lowest informativeness in the CC *is*-contraction data set are *one*, *mum*, *dad*, and *thing*. These have a far higher rate of contractions (83.7%) than the average for nonpronouns. Some authentic examples appear in 34.

- (34) a. and my poor **Mum's** here going oh I wish I was there
 b. and I said come quick come quick . **Dad's** at home and he's a hell of a mess
 c. **one's** a um . a raving . feminist an **one's** a chauvinist
 d. I've got [a] friend that has three cats and **one's** a really spiteful cat .
 e. liturgy that they all join in on . and the whole **thing's** sung
 f. I wonder if that that kind of **thing's** like hereditary

These simple descriptive statistics support this crucial consequence of the hybrid theory: that *is* contraction with nonpronoun hosts should depend on the usage probabilities of specific lexical hosts. But while these data points are suggestive, what is needed to test the prediction is a statistical model that controls for other possible predictors of contraction. After all, there are many hosts in the data set, and the literature on contraction has identified many contributors to *is* contraction other than informativeness (see below). To this end, a multiple logistic regression model was fit to the nonpronoun host data ($n = 1,398$) annotated 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.

8.1. VARIABLES.

INFORMATIVENESS. The main variable of interest, the informativeness of the nonpronoun host before *is/s*, is calculated as described above.

HOST PHRASE WORD COUNT. Host phrase word count (WC) is one of the best predictors of contraction (Frank & Jaeger 2008, Bresnan & Spencer 2012, MacKenzie 2012, 2013,

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.³⁷ 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.³⁸ Table 5 provides authentic examples. Table 6 shows the relation of word counts (excluding pronoun hosts and counting space-separated character strings by an R script) to contractions in the data. All host phrases were manually identified.

HOST PHRASES (bolded)	WORD COUNT
but now work 's just so busy ...	WC = 1
the work 's so much harder	WC = 2
all this blood 's pouring out the side of my head	WC = 3
some of the work is a bit tedious	WC = 4

TABLE 5. Host phrase word count.

HOST PHRASE WC	1	2	3	4+
total instances	543	500	182	143
proportion contracted	0.74	0.55	0.43	0.18

TABLE 6. Proportion contracted by host phrase word count.

YEAR OF BIRTH. With nonpronoun 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. 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.

YEAR OF BIRTH	[1926,1961)	[1961,1987]
proportion contracted	0.50	0.61

TABLE 7. Proportion contracted by speaker year of birth.

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

CLASS	NONPROFESSIONAL	PROFESSIONAL
proportion contracted	0.65	0.49

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'. Successive instances of *is*/*'s* are from the same speaker, are collapsed across the copula/auxiliary types (see below), and include all previous contractions, including those with pronoun hosts.

PREVIOUS INSTANCE	<i>'s</i>	<i>is</i>	NONE
proportion contracted	0.599	0.342	0.510

TABLE 9. Proportion contracted by previous occurrence of *is*/*'s*.

is TYPE. The copular and auxiliary uses of *is*/*'s* were manually identified. Those instances of *is*/*'s* in construction with a participial form of the verb are defined as 'auxil-

³⁷ See Szmrecsányi 2004 on operationalizing syntactic complexity.

³⁸ However, 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 data set.

iaries', 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.³⁹

<i>is</i> TYPE	AUX	COP
proportion contracted	0.635	0.548

TABLE 10. Proportion contracted by *is* type.

8.2. 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 host phrase were tested as alternatives to WC for another project: (i) total metrical feet (Sternberg et al. 1978, Sternberg et al. 1988); (ii) 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); (iii) cumulative stress from manual annotation of perceived stress values, with and without transformation to a grid format (Lieberman & Prince 1977); and (iv) cumulative stress based on manually corrected automatic annotation of theoretical stress values, transformed to grid formats. (i) and (iv) were automatically annotated using software developed by Anttila et al. (2020). WC substantially improves the model fit compared to alternatives (i) and (iv), while (ii) and (iii) are both competitive with WC. WC is retained here as a convenient proxy pending further research.

8.3. THE FITTED MODEL. Because speaker identity is a source of unknown dependencies in the data, a multiple logistic regression 'working independence' model (Harrell 2001) was constructed from these variables, with the numerical variables standardized.⁴⁰ After the model was fitted to the data, it was corrected for intraspeaker correlations by bootstrap cluster sampling with replacement using the `bootcov()` function of Harrell 2018. The resulting parameter values are shown in the final fitted model in Table 11.

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

³⁹ 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.

⁴⁰ 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. (2007:83) describe cluster resampling in this way (emphasis in original):

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.*

$$\text{Prob}\{\text{Contracted} = 1\} = \frac{1}{1 + e^{-\lambda\beta}}, \text{ where}$$

$$\lambda\beta =$$

$$\begin{aligned} & 0.8804 \\ & -0.4741 \times -\log_2 P(\text{host}|\text{verb}) \\ & -0.9868 \times [\text{previous instance} = \text{is}] \\ & -0.2177 \times [\text{previous instance} = \text{none}] \\ & -1.0068 \times \text{host phrase WC} \\ & -0.7060 \times [\text{class} = \text{P}] \\ & -0.5370 \times [\text{is type} = \text{cop}] \\ & +0.4515 \times [\text{year of birth} = [1961, 1987]] \\ & \text{and } [c] = 1 \text{ if subject is in group } c, 0 \text{ otherwise} \end{aligned}$$

TABLE 11. Model of Canterbury Corpus variable *is*-contraction data with nonpronoun hosts.

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.⁴¹ The noncategorical predictors $-\log_2 P(\text{host}|\text{verb})$ (INFORMATIVENESS of the host given the verb) and HOST PHRASE WC have scalar values that 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.⁴² Partial effects of the model are plotted in Figure 12. 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.

This finding was replicated on the nonpronoun-host data from the BC. The predictors are the same except for age and class, which were unavailable or unrelated to contraction in this data set. Modeling and validation by the same methods as before showed a reliable effect of informativeness of the host on contraction.⁴³

Barth and Kapatsinski (2017:40–41) conducted a multimodel analysis of *is*/*ʒ* contractions with nonpronoun hosts in a smaller data set of spoken language from the Cor-

⁴¹ The three-valued predictor for PREVIOUS INSTANCE is decomposed into two binary two-valued predictors: full *is* vs. *ʒ*, and no previous instance vs. *ʒ*.

⁴² Validation of the model found that a proportion greater than 0.95 of averaged observed minus expected values in thirty-five bins are within 2 standard errors (see Gelman and Su’s (2018) `binnedplot()` function); all predictors have low multicollinearity (condition number $c < 5$, $\text{vif} < 1.1$); average concordance is $C > 0.758$ under ten-fold crossvalidation with bias correction for speaker clusters in each fold—an ‘optimism’ of < 0.01 .

⁴³ For the BC replications, a data set of variable *is* contractions was extracted and annotated following a similar method to that of Bresnan and Spencer (2012) and J. Spencer (2014), who already show an effect on contraction of the log or negative log conditional probability of nonpronoun hosts given *is*/*ʒ* in data collected from the BC. The replication data set in the present study was constructed independently of the data sets described in those studies and encompasses a greater range of host phrase lengths.

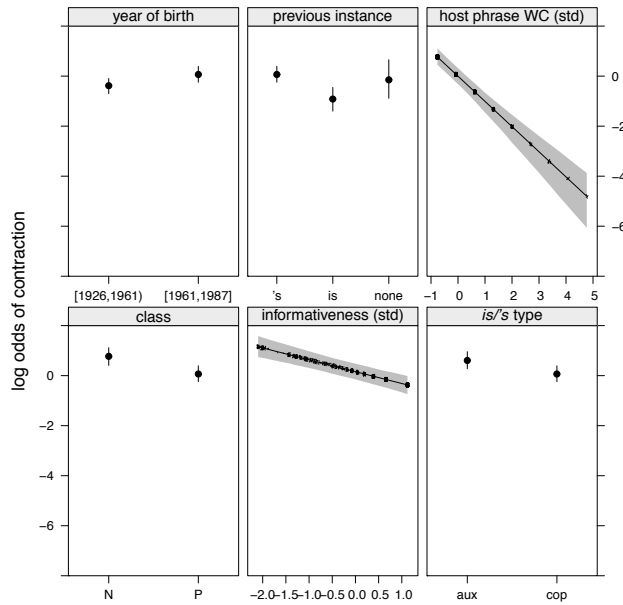


FIGURE 12. Partial effects of the model in Table 11. Each panel shows the effect of one predictor when all of the others are held constant. 95% confidence bands are from the bootstrapped cluster resampling of speakers. The vertical ticks on the plot lines of the numerical covariates (host phrase WC and host informativeness) show the data densities along the predictor scales.

pus 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. 7).

In sum, this prediction of a 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): usage probabilities affect not only the contractions of restrictive auxiliaries with their pronoun hosts and morphophonological fusions; they also affect in the same way the contractions of the most unrestricted auxiliary *is* with noun hosts.

9. BETWEEN RESTRICTIVE AND NONRESTRICTIVE. As §7 points out, the restrictive auxiliaries’ syntactic restrictions to subject pronoun and interrogative proform hosts single out the syntactic positions that have the highest token frequencies of cooccurrence with the auxiliaries (cf. 23 and n. 31). The formal descriptions of these syntactic restrictions in the lexical entries of auxiliaries in §§5 and 6 can then be regarded as describing grammaticalizations of distributional usage patterns. A closer examination reveals that, as one might expect from the grammaticalization of usage patterns, the line between restrictive and nonrestrictive auxiliaries is not a binary categorical classification as implied in §4.

Although the restrictive auxiliary *’ve* overwhelmingly occurs with subject pronouns and interrogative proforms, with low probability it does contract with some host nouns, such as example 35 (Barron 1998:247, n. 13) and example 36 from the BC.

(35) The **BBC***’ve* reported ... [bi:bi’si:v]

(36) ... all their life **people***’ve* been saying ... [’pi:pɪv]

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.

(41) Lexical entry for *Bligh'd*

$$x [\dots d] \leftarrow \begin{array}{c} X \\ \downarrow = \downarrow \\ (\downarrow \text{MOOD}) = \text{COND} \\ (\downarrow \text{SUBJ}) = {}_c \downarrow \end{array} \begin{array}{c} I \\ \downarrow = \downarrow \\ (\downarrow \text{MOOD}) = \text{COND} \\ (\downarrow \text{SUBJ}) = {}_c \downarrow \end{array}$$

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* *[soʊd] for *So would Ann*, as well as accounting for **How'd it have happened?* in 37e.

For the present author 'd is even less restrictive, allowing contractions not only with a subject, as in 41, but also with an adjacent dependent of the subject: witness *family'd* in 26. The greater degree of contraction is permitted by the lexical entry in 42.⁴⁶

(42) Lexical entry for *family'd*

$$x [\dots d] \leftarrow \begin{array}{c} X \\ \downarrow = \downarrow \\ (\downarrow \text{MOOD}) = \text{COND} \\ (\downarrow \text{SUBJ GF}^*) = {}_c \downarrow \end{array} \begin{array}{c} I \\ \downarrow = \downarrow \\ (\downarrow \text{MOOD}) = \text{COND} \\ (\downarrow \text{SUBJ}) = {}_c \downarrow \end{array}$$

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*.

It is plausible that usage probabilities underlie these specific differences in syntactic distribution, shaping the synchronic grammar of auxiliaries as they have their diachronic development (Bybee 2010). But with the relative paucity of 'd contractions in corpora, and the infrequency of long host phrases in spoken language in general (n. 38), the necessary research would probably require experimental methods beyond the scope of the present study.⁴⁷

The formal syntactic analyses illustrated above also suggest a path by which auxiliaries can change from one type to another: it is by a kind of 'syntactic bleaching' in which relational features are gradually lost, initially by becoming optional, reflecting variable restrictiveness, and eventually by dropping the feature option altogether, shifting not the meaning but the distribution of the auxiliary.⁴⁸ The auxiliaries in the respective entries 40 and 41 for the British and American varieties differ by the loss of the feature specifying a pronoun host. A fully unrestricted 'd parallel to the unrestricted 's in 30 would differ from both 41 and 42 by the loss of the feature constraining the host to be the subject of the auxiliary. A rich lexical syntactic literature on the development of agreement markers from pronoun clitics in multiple languages (see Bresnan et al. 2015:Ch. 8 and references there) shows that this kind of feature optionality and loss is a natural progression that is well captured by the relational specifications of the formal grammar.

10. I dunno PARALLELS AND IMPLICATIONS. The formal theory of lexically shared host + auxiliary contractions extends further into the larger domain of multiword ex-

⁴⁶ 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.

⁴⁷ Experimental psycholinguistic studies have found phrase frequency effects on production (e.g. Bannard & Matthews 2008, Janssen & Barber 2012, Arnon & Cohen Priva 2013, Shao et al. 2019); see Jacobs et al. 2016 for a review of frequency effects of word sequences in production and comprehension.

⁴⁸ The grammaticalization literature characterizes semantic change as involving, in part, a loss or shift of lexical meaning, sometimes termed 'semantic bleaching', although semantic shifting may be a more accurate depiction (Hopper & Traugott 2003).

pressions, such as Bybee and Scheibman's (1999) study of *I don't know* discussed in §1. The formal analysis brings out parallels between this multiword expression and the grammar of tensed auxiliary contractions.

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

Second, Bybee and 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.

- (43) a. I'll [aɪ/aɪ] certainly come.
 I [aɪ/*ɑ] certainly'll [əɪ/*ɪ] come.
 b. They're [ðeɪ/ðeɪ] certainly expensive.
 They [ðeɪ/*ðeɪ] certainly're [əɪ/*ɪ] expensive.

Third, *don't* reduction fails with a conjoined pronoun *I* and with a lexical subject (Kaisse 1985, Scheibman 2000), as 44a,b illustrate. (Following Scheibman 2000, the orthographic representation of reduced *I don't know* as *I dunno* is used here.)

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

The same syntactic restrictions characterize the restrictive contractions, as already seen in examples 23a–c.

The illustrative lexical entries in 45–47 are sufficient to capture all three properties of parallelism between contraction and *I dunno* reduction:⁴⁹ (i) the dependence on the specific pronoun *I* for the pronunciation of *don't* as [rɔ̃], (ii) the required adjacency of *I* and *don't* for this reduced pronunciation, and (iii) the syntactic restrictions against a conjoined subject with *I* (44a) and against a lexical noun phrase subject (44b).

- (45) Lexical entry for *don't*
- | | | |
|------------------|---|--------------------|
| don't [dɔ̃t/dɔ̃] | ← | I |
| | | (↓ TENSE) = PRES |
| | | (↓ POLARITY) = NEG |
| | | ¬(↓ SUBJ PERS) = 3 |
| | | ↓ = ↓ |

- (46) Lexical entry for *I don't*
- | | | | | |
|-----------------|---|------------------|--|---------------------------|
| I don't [aɪrɔ̃] | ← | D | | I |
| | | (↓ PRED) = 'PRO' | | (↓ TENSE) = PRES |
| | | (↓ PERS) = 1 | | (↓ POLARITY) = NEG |
| | | (↓ NUM) = SG | | ¬(↓ SUBJ PERS) = 3 |
| | | ↓ = ↓ | | (↓ SUBJ) = _c ↓ |

⁴⁹ Zwicky and Pullum (1983) provide evidence that *n't* is an inflectional affix; see also Huddleston & Pullum 2002.

11. CONCLUDING DISCUSSION. The present study provides a high-level description of how a hybrid of formal grammar and the usage-based mental lexicon could explain the combined findings on tensed auxiliary contractions in English from both usage-based and formal lines of research. There are other architectures for exemplar models of syntax that might be adopted. The dual-route multilevel exemplar model of Walsh et al. (2010) is noteworthy. Their innovative contribution is to formalize the relations between CONSTITUENTS and UNITS explicitly at both the phonetic and syntactic levels. For example, segments are constituents of syllable units, and words are constituents of phrase or sentence units. These are stored in memory and categorized into clouds of exemplars according to their similarity to existing exemplars. The architecture of their model employs two routes from every input to the output, setting up a competition between a submodel that directly selects the output as a unit exemplar and a submodel that assembles exemplar constituents into an output: the unit submodel wins if the unit exemplar receives activation above a threshold. Although Walsh et al. (2010) discuss the goal of modeling phonetically detailed phrases stored in memory (e.g. Hay & Bresnan 2006) and Schütze et al. (2007) in a related paper simulate the grammaticalization of *going to* (Bybee 2006), the Bybee-Pierrehumbert model adopted here more directly connects with the data of the present study.

Particularly interesting is that Walsh et al.'s (2010) models do not make any use at all of representational labels from formal grammar, whether phonological or syntactic. Their syntactic model achieves impressive results in learning grammaticality judgments of simple sentences (for example, *I like tea* vs. **I tea like*) from a purely quantitative distributional analysis of words in a corpus of child-directed speech to children of ages two to three years.⁵⁰ How this approach could extend to the complexities of adult grammatical knowledge, which include nonlocal word-order dependencies (e.g. *Is it tea that you said that you like?*), remains to be seen. At bottom, all syntactic categories are distributional: 'The similar syntactic behavior of two nouns like *coin* and *hen* is not directly apparent from their pronunciation or semantics. But an exemplar-theoretic account of syntactic behavior requires a similarity measure where *coin* and *hen* are similar' (Walsh et al. 2010:561–62). Although relational features like SUBJECT of course involve a much higher level of abstraction than sequential parts of speech (Bresnan et al. 2015), Walsh et al.'s (2010) multilevel exemplar model is not fundamentally incompatible with the hybrid model sketched here.⁵¹

One other computationally explicit syntactic exemplar model is Bod's (1998, 2006, 2009) data-oriented parsing (DOP) model, in which the corpus is the grammar. Bod and Kaplan (1998) and Bod (2006) show how the DOP model employing LFG c-structure to f-structure mappings can achieve productivity by parsing unseen data through structural

⁵⁰ Building on a machine-learning approach to part-of-speech tagging (Schütze 1995), their model assigns each word two vectors, one consisting of the probabilities of all of its left-context words in the corpus and the other those of its right-context words, computed using relative frequencies that correspond to the maximum likelihood estimate for each probability. A word's similarity to exemplar words is measured by the sum of the cosines of these vectors (the same similarity measure used at the phonetic level in their syllable production model). In a simulation, Walsh et al. (2010) demonstrate that their distributional method of assigning fine-grained and gradient parts of speech to words performs better than category-based representations in judging the grammaticality of word-order permutations of simple sentences.

⁵¹ In their conclusion Walsh et al. (2010:575) suggest that their model could be the basis for hybrid models of later stages of language development, with exemplar clouds linked to more abstract layers of representation, referring explicitly to the informal hybrid model of acquisition of Abbot-Smith and Tomasello (2006).

analogy. In a very interesting later article, Bod (2009) shows how an unsupervised parser of data from the Eve corpus (Brown 1973) in the CHILDES database (MacWhinney 2000) can learn auxiliary inversion (a paradigm example of the seeming need for innate syntactic categories to overcome the ‘poverty of the stimulus’ in language learning). The Pierrehumbert 2001 model adopted here provides a shorter and clearer path from the morphophonological data to the syntax of tensed auxiliary contraction.

The present study also makes an empirical contribution specific to LFG’s theory of formal grammar in demonstrating the explanatory value of multiword lexically shared expressions (as does Broadwell 2007:n. 3).⁵² CONSTRUCTION GRAMMAR (Fillmore et al. 1988, Goldberg 1995, 2006, Croft 2001) already allows lexical representations of multiword expressions as constructions, as does the DOP model. The formal grammar of the present study shares a number of linguistic features with construction grammar, including the storage of lexically specific constructions (for example, Figs. 4 and 13) and lexical schemata for productive constructions (30). Where construction grammar aims to derive semantic distinctions among lexical words from their constructional contexts rather than from multiple lexical entries, the present study focuses on the usage-based lexicalization of syntactic fragments. There is no reason why the present framework could not be extended to other areas of grammar where usage affects the semantics and pragmatics of multiword expressions.

Bybee’s conception of constructions in several works appears to challenge the role of constituent structure as a systematic level of representation. Bybee and Scheibman (1999) suggest that the erosion of internal constituent structure boundaries is associated with the phonetic fusions of frequently cooccurring words. While this erosion demonstrably occurs with frozen contractions in expressions like *whosie-whatsit*, *howsit/howzit* (n. 17), the evidence in §§4–5 shows that restrictive auxiliaries retain their constituent structure despite lexically specific phonetic fusions with their hosts. These contractions are intermediate between frozen lexicalizations and full syntactic phrases: they show phonetic compression and fusion, but retain syntactic life.

In a subsequent work, Bybee (2002) appears to argue against hierarchical constituent structure altogether, at one point proposing (p. 130): ‘Constituents of the type proposed for generative grammar which are described by phrase structure trees do not exist. Instead, units of language (words or morphemes) are combined into chunks as a result of frequent repetition’. Based on the evidence that contractions like *you’re* and similar units are chunks that overlap with c-structure constituents like NP and VP rather than nest hierarchically within them, one might suppose that these chunks undermine the concept of hierarchical c-structure trees.⁵³

In contrast, the present study shows that *you’re* can both be a lexical-syntactic unit or ‘chunk’ and share a common hierarchical c-structure with *you are*. The same is true of other common fragments such as *in the middle of* (Tremblay & Baayen 2010). The present study empirically confirms that tensed auxiliary contractions are lexicalized

⁵² Lowe (2016) proposes a dual-source analysis of the English \mathfrak{s} genitive similar in spirit to the analysis of tensed \mathfrak{s} of the present study: he assumes genitive \mathfrak{s} is a clitic except in cases where lexical sharing with the host is motivated. However, his version of lexical sharing differs. Wescoat’s λ mapping from c-structure to l(exical)-structure is a homomorphism, which entails the adjacency, linear order, and lexical integrity of the atomic components of lexical sharing (n. 35; Wescoat 2005, 2009). Lowe’s π mapping is an inverse of λ and hence is a relation, not a function; it requires separately stipulating the linear order of atomic components of his shared entries, as well as their lexical integrity.

⁵³ Whether Bybee herself would make this argument is difficult to determine from her informal depictions of syntactic abstractions and constructions.

chunks in Bybee's sense, but shows that they coexist as syntactic fragments from a system of explicit mappings between surface constituent structures and abstract grammatical dependencies.⁵⁴

The main contribution of the present study has been to describe and provide novel evidence for a hybrid model of tensed auxiliary contractions. The novel evidence includes (i) a synthesis of morphophonological, prosodic, and syntactic findings on tensed auxiliary contraction from both formal and usage-based research, (ii) a corpus study confirming core properties of the hybrid model, (iii) a formal analysis of the grammaticalization of host-auxiliary restrictions from their distributional usage patterns, and (iv) the extension of the formal grammar of auxiliary contraction to a multiword expression of classic usage-based grammar (Bybee & Scheibman 1999), revealing surprising parallels. These results show the empirical and theoretical value of combining formal and usage-based data and methods into a more explanatory shared framework.

REFERENCES

- ABBOT-SMITH, KIRSTEN, and MICHAEL TOMASELLO. 2006. Exemplar-learning and schematization in a usage-based account of syntactic acquisition. *The Linguistic Review* 23. 275–90. DOI: 10.1515/TLR.2006.011.
- ABERCROMBIE, DAVID. 1961. Syllable quantity and enclitics in English. In *honour of Daniel Jones: Papers contributed on the occasion of his eightieth birthday, 12 September 1961*, ed. by David Abercrombie, Dennis B. Fry, Peter A. D. MacCarthy, Norman C. Scott, and J. L. M. Trim, 216–22. London: Longmans, Green, and Co.
- ALSINA, ALEX. 2010. The Catalan definite article as lexical sharing. *Proceedings of the LFG '10 Conference*, 5–25. Online: <http://web.stanford.edu/group/csli/publications/csli/publications/LFG/15/papers/lfg10alsina.pdf>.
- AMBRIDGE, BEN. 2020. Against stored abstractions: A radical exemplar model of language acquisition. *First Language* 40.509–59. DOI: 10.1177/0142723719869731.
- ANDERSON, STEPHEN R. 2005. *Aspects of the theory of clitics*. Oxford: Oxford University Press.
- ANDERSON, STEPHEN R. 2008. English reduced auxiliaries really are simple clitics. *Lingue e linguaggio* 7.169–86. DOI: 10.1418/28094.
- ANTILA, ARTO. 2017. Stress, phrasing, and auxiliary contraction in English. *The morphosyntax-phonology connection: Locality and directionality at the interface*, ed. by Vera Griбанова and Stephanie Shih, 143–70. Oxford: Oxford University Press. DOI: 10.1093/acprof:oso/9780190210304.003.0006.
- ANTILA, ARTO; TIMOTHY DOZAT; DANIEL GALBRAITH; and NAOMI SHAPIRO. 2020. Sentence stress in presidential speeches. *Prosody in syntactic encoding* (Linguistische Arbeiten 573), ed. by Gerrit Kentner and Joost Kremers, 17–50. Berlin: De Gruyter. DOI: 10.1515/9783110650532-002.
- ARNON, INBAL, and URIEL COHEN PRIVA. 2013. More than words: The effect of multi-word frequency and constituency on phonetic duration. *Language and Speech* 56.349–71. DOI: 10.1177/0023830913484891.
- BAKER, CARL L. 1971. Stress level and auxiliary behavior in English. *Linguistic Inquiry* 2. 167–81. Online: <https://www.jstor.org/stable/4177623>.
- BANNARD, COLIN, and DANIELLE MATTHEWS. 2008. Stored word sequences in language learning: The effect of familiarity on children's repetition of four-word combinations. *Psychological Science* 19.241–48. DOI: 10.1111/j.1467-9280.2008.02075.x.
- BARRON, JULIA. 1998. *Have* contraction: Explaining 'trace effects' in a theory without movement. *Linguistics* 36.223–51. DOI: 10.1515/ling.1998.36.2.223.
- BARTH, DANIELLE. 2011. Modeling reduction of *is*, *am* and *are* in grammaticalized constructions. *Quantitative Issues in Theoretical Linguistics (QITL)* 4.13–15. Online: <https://edoc.hu-berlin.de/handle/18452/2016>.

⁵⁴ See Bresnan 2001 or Bresnan et al. 2015 on the fragmentability of language in the LFG formal architecture.

- BARTH, DANIELLE. 2019. Effects of average and specific context probability on reduction of function words *BE* and *HAVE*. *Linguistics Vanguard* 5:20180055. DOI: 10.1515/lingvan-2018-0055.
- BARTH, DANIELLE, and 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* 13.203–60. DOI: 10.1515/cllt-2014-0022.
- BENDER, EMILY M., and IVAN A. SAG. 2001. Incorporating contracted auxiliaries in English. *Grammatical interfaces in HPSG*, ed. by Ronnie Cann, Claire Grover, and Philip Miller, 1–15. Stanford, CA: CSLI Publications.
- BOD, RENS. 1998. *Beyond grammar: An experience-based theory of language*. Stanford, CA: CSLI Publications.
- BOD, RENS. 2006. Exemplar-based syntax: How to get productivity from examples. *The Linguistic Review* 23.291–320. DOI: 10.1515/TLR.2006.012.
- BOD, RENS. 2009. From exemplar to grammar: A probabilistic analogy-based model of language learning. *Cognitive Science* 33.752–93. DOI: 10.1111/j.1551-6709.2009.01031.x.
- BOD, RENS, and RONALD KAPLAN. 1998. A probabilistic corpus-driven model for lexical-functional analysis. *Proceedings of the 17th International Conference on Computational Linguistics*, vol. 1, 145–51. Online: <https://www.aclweb.org/anthology/C98-1022>.
- BÖRJARS, KERSTI; RACHEL NORDLINGER; and LOUISA SADLER. 2019. *Lexical-functional grammar: An introduction*. Cambridge: Cambridge University Press. DOI: 10.1017/9781316756584.
- BRESNAN, JOAN. 1971. Contraction and the transformational cycle in English. Cambridge, MA: MIT, ms. [Reproduced by the Indiana University Linguistics Club, 1978.]
- BRESNAN, JOAN. 1973. Syntax of the comparative clause construction in English. *Linguistic Inquiry* 4.275–343. Online: <https://www.jstor.org/stable/4177775>.
- BRESNAN, JOAN. 2001. *Lexical-functional syntax*. Oxford: Blackwell.
- BRESNAN, JOAN. 2016. Linguistics: The garden and the bush. *Computational Linguistics* 42. 599–617. DOI: 10.1162/COLI_a_00260.
- BRESNAN, JOAN; ASH ASUDEH; IDA TOIVONEN; and STEPHEN WECHSLER. 2015. *Lexical-functional syntax*. 2nd edn. (Blackwell textbooks in linguistics.) Oxford: Wiley-Blackwell.
- BRESNAN, JOAN; ANNA CUENI; TATIANA NIKITINA; and R. HARALD BAAYEN. 2007. Predicting the dative alternation. *Cognitive foundations of interpretation*, ed. by Gerlof Bouma, Irene Krämer, and Joost Zwarts, 69–94. Amsterdam: Royal Netherlands Academy of Science.
- BRESNAN, JOAN, and JESSICA SPENCER. 2012. Frequency effects in spoken syntax: *Have* and *be* contraction. Invited paper presented at the symposium on New Ways of Analyzing Syntactic Variation, Radboud University Nijmegen, 15–17 November 2012.
- BROADWELL, GEORGE AARON. 2007. Lexical sharing and non-projecting words: The syntax of Zapotec adjectives. *Proceedings of the LFG '07 Conference*, 87–106. Online: <http://web.stanford.edu/group/cslipublications/cslipublications/LFG/12/papers/lfg07broadwell.pdf>.
- BROADWELL, GEORGE AARON. 2008. Turkish suspended affixation is lexical sharing. *Proceedings of the LFG '08 Conference*, 198–213. Online: <http://web.stanford.edu/group/cslipublications/cslipublications/LFG/13/papers/lfg08broadwell.pdf>.
- BROWN, ROGER. 1973. *A first language: The early stages*. Cambridge, MA: Harvard University Press.
- BRYLSBAERT, MARC; MICHAËL STEVENS; PAWEŁ MANDERA; and EMMANUEL KEULEERS. 2016. The impact of word prevalence on lexical decision times: Evidence from the Dutch Lexicon Project 2. *Journal of Experimental Psychology: Human Perception and Performance* 42.441–58. DOI: 10.1037/xhp0000159.
- BYBEE, JOAN. 1985. *Morphology: A study of the relation between meaning and form*. Amsterdam: John Benjamins.
- BYBEE, JOAN. 2000. The phonology of the lexicon: Evidence from lexical diffusion. *Usage-based models of language*, ed. by Michael Barlow and Suzanne Kemmer, 65–85. Stanford, CA: CSLI Publications.
- BYBEE, JOAN. 2001. *Phonology and language use*. Cambridge: Cambridge University Press.
- BYBEE, JOAN. 2002. Sequentiality as the basis of constituent structure. *The evolution of language from pre-language*, ed. by T. Givón and Bertram F. Malle, 109–32. Amsterdam: John Benjamins.

- BYBEE, JOAN. 2006. From usage to grammar: The mind's response to repetition. *Language* 82.711–33. DOI: 10.1353/lan.2006.0186.
- BYBEE, JOAN. 2010. *Language, usage, and cognition*. Cambridge: Cambridge University Press.
- BYBEE, JOAN, and PAUL J. HOPPER (eds.) 2001. *Frequency and the emergence of linguistic structure*. Amsterdam: John Benjamins.
- BYBEE, JOAN, and JAMES L. MCCLELLAND. 2005. Alternatives to the combinatorial paradigm of linguistic theory based on domain general principles of human cognition. *The Linguistic Review* 22.381–410. DOI: 10.1515/tlir.2005.22.2-4.381.
- BYBEE, JOAN, and JOANNE SCHEIBMAN. 1999. The effect of usage on degrees of constituency: The reduction of *don't* in English. *Linguistics* 37.575–96. DOI: 10.1515/Ling.37.4.575.
- CROFT, WILLIAM. 2001. *Radical construction grammar: Syntactic theory in typological perspective*. Oxford: Oxford University Press.
- DALRYMPLE, MARY; JOHN J. LOWE; and LOUISE MYCOCK. 2019. *The Oxford reference guide to lexical functional grammar*. Oxford: Oxford University Press.
- DAVIES, MARK. 2008–. *Corpus of Contemporary American English: 400+ million words, 1990–present*. Online: <https://www.english-corpora.org/coca/>.
- DIESSEL, HOLGER. 2007. Frequency effects in language acquisition, language use, and diachronic change. *New Ideas in Psychology* 25.108–27. DOI: 10.1016/j.newideapsych.2007.02.002.
- DIXON, ROBERT M. W. 1977. Semantic neutralization for phonological reasons. *Linguistic Inquiry* 8.599–602. Online: <https://www.jstor.org/stable/4178004>.
- ERNESTUS, MIRJAM. 2014. Acoustic reduction and the roles of abstractions and exemplars in speech processing. *Lingua* 142.27–41. DOI: 10.1016/j.lingua.2012.12.006.
- FILLMORE, CHARLES J.; PAUL KAY; and MARY CATHERINE O'CONNOR. 1988. Regularity and idiomatity in grammatical constructions: The case of *let alone*. *Language* 64.501–38. DOI: 10.2307/414531.
- FRANK, AUSTIN, and T. FLORIAN JAEGER. 2008. Speaking rationally: Uniform information density as an optimal strategy for language production. *Proceedings of the 30th annual meeting of the Cognitive Science Society (CogSci '08)*, 939–44.
- GELMAN, ANDREW, and YU-SUNG SU. 2018. arm: Data analysis using regression and multi-level/hierarchical models. R package version 1.10-1. Online: <https://cran.r-project.org/package=arm>.
- GERMAN, JAMES; JANET BRECKENRIDGE PIERREHUMBERT; and STEFAN KAUFMANN. 2006. Evidence for phonological constraints on nuclear accent placement. *Language* 82.151–68. DOI: 10.1353/lan.2006.0035.
- GODFREY, JOHN J., and EDWARD HOLLIMAN. 1997. Switchboard-1 release 2. Philadelphia: Linguistic Data Consortium.
- GOLDBERG, ADELE E. 1995. *Constructions: A construction grammar approach to argument structure*. Chicago: University of Chicago Press.
- GOLDBERG, ADELE E. 2006. *Constructions at work: The nature of generalization in language*. Oxford: Oxford University Press.
- GORDON, ELIZABETH; LYLE CAMPBELL; JENNIFER HAY; MARGARET MACLAGAN; ANDREA SUDBURY; and PETER TRUDGILL. 2004. *New Zealand English: Its origins and evolution*. Cambridge: Cambridge University Press.
- HARMON, ZARA, and VSEVOLOD KAPATSINSKI. 2017. Putting old tools to novel uses: The role of form accessibility in semantic extension. *Cognitive Psychology* 98.22–44. DOI: 10.1016/j.cogpsych.2017.08.002.
- HARRELL, FRANK E., JR. 2001. *Regression modeling strategies: With applications to linear models, logistic regression, and survival analysis*. Dordrecht: Springer.
- HARRELL, FRANK E., JR. 2018. rms: Regression modeling strategies. R package version 5.1-2. Online: <https://cran.r-project.org/package=rms>.
- HAY, JENNIFER, and JOAN BRESNAN. 2006. Spoken syntax: The phonetics of *giving a hand* in New Zealand English. *The Linguistic Review* 23.321–49. DOI: 10.1515/TLR.2006.013.
- HOPPER, PAUL J., and ELIZABETH CLOSS TRAUOGOTT. 2003. *Grammaticalization*. 2nd edn. Cambridge: Cambridge University Press.
- HUDDLESTON, RODNEY, and GEOFFREY K. PULLUM. 2002. *The Cambridge grammar of the English language*. Cambridge: Cambridge University Press.

- INKELAS, SHARON. 1991. *Prosodic constituency in the lexicon*. New York: Garland.
- INKELAS, SHARON, and DRAGA ZEC. 1993. Auxiliary reduction without empty categories: A prosodic account. *Working Papers of the Cornell Phonetics Laboratory* 8.205–53. DOI: 10.5281/zenodo.3735001.
- ITO, JUNKO, and ARMIN MESTER. 2018. Matching light elements. *A reasonable way to proceed: Essays in honor of Jim McCloskey*, ed. by Jason Merchant, Line Mikkelsen, Deniz Rudin, and Kelsey Sasaki, 168–90. Santa Cruz: University of California, Santa Cruz. Online: <https://escholarship.org/uc/item/7z29n70x>.
- JACOBS, CASSANDRA L.; GARY S. DELL; AARON S. BENJAMIN; and COLIN BANNARD. 2016. Part and whole linguistic experience affect recognition memory for multiword sequences. *Journal of Memory and Language* 87.38–58. DOI: 10.1016/j.jml.2015.11.001.
- JANSSEN, NIELS, and HORACIO A. BARBER. 2012. Phrase frequency effects in language production. *PLoS ONE* 7:e3320. DOI: 10.1371/journal.pone.0033202.
- KAISSE, ELLEN M. 1979. Auxiliary reduction and the derivation of pseudoclefts. *Linguistic Inquiry* 10.707–10. Online: <https://www.jstor.org/stable/4178139>.
- KAISSE, ELLEN M. 1983. The syntax of auxiliary reduction in English. *Language* 59.93–122. DOI: 10.2307/414062.
- KAISSE, ELLEN M. 1985. *Connected speech: The interaction of syntax and phonology*. New York: Academic Press.
- KIESLING, SCOTT; LAURA DILLEY; and WILLIAM D. RAYMOND. 2006. The Variation in Conversation (ViC) project: Creation of the Buckeye Corpus of conversational speech. Columbus: Department of Psychology, The Ohio State University. Online: <https://buckeyecorpus.osu.edu/BuckeyeCorpusmanual.pdf>.
- KING, HAROLD V. 1970. On blocking the rules for contraction in English. *Linguistic Inquiry* 1.134–36. Online: <https://www.jstor.org/stable/4177538>.
- KLAVANS, JUDITH L. 1985. The independence of syntax and phonology in cliticization. *Language* 61.95–120. DOI: 10.2307/413422.
- KRUG, MANFRED. 1998. String frequency: A cognitive motivating factor in coalescence, language processing, and linguistic change. *Journal of English Linguistics* 26.286–320. DOI: 10.1177/007542429802600402.
- LABOV, WILLIAM. 1969. Contraction, deletion, and inherent variability of the English copula. *Language* 45.715–62. DOI: 10.2307/412333.
- LEVELT, WILLEM J. M.; ARDI ROELOFS; and ANTIJE S. MEYER. 1999. A theory of lexical access in speech production. *Behavioral and Brain Sciences* 22.1–38. DOI: 10.1017/S0140525X99001776.
- LIBERMAN, MARK, and ALAN PRINCE. 1977. On stress and linguistic rhythm. *Linguistic Inquiry* 8.249–336. Online: <https://www.jstor.org/stable/4177987>.
- LOWE, JOHN J. 2016. English possessive 's: Clitic and affix. *Natural Language and Linguistic Theory* 34.157–95. DOI: 10.1007/s11049-015-9300-1.
- MACKENZIE, LAUREL E. 2012. *Locating variation above the phonology*. Philadelphia: University of Pennsylvania dissertation.
- MACKENZIE, LAUREL E. 2013. Variation in English auxiliary realization: A new take on contraction. *Language Variation and Change* 25.17–41. DOI: 10.1017/S0954394512000257.
- MACWHINNEY, BRIAN. 2000. *The CHILDES project: Tools for analyzing talk: The database*, vol. 2. London: Psychology Press.
- MCELHINNY, BONNIE S. 1993. Copula and auxiliary contraction in the speech of white Americans. *American Speech* 68.371–99. DOI: 10.2307/455773.
- MEL'ČUK, IGOR. 1988. *Dependency syntax: Theory and practice*. Albany: SUNY Press.
- NATHAN, GEOFFREY S. 1981. What's these facts about? *Linguistic Inquiry* 12.151–53. Online: <https://www.jstor.org/stable/4178211>.
- PHILLIPS, BETTY S. 1984. Word frequency and the actuation of sound change. *Language* 60.320–42. DOI: 10.2307/413643.
- PIANTADOSI, STEVEN T.; HARRY TILY; and EDWARD GIBSON. 2011. Word lengths are optimized for efficient communication. *Proceedings of the National Academy of Sciences* 108.3526–29. DOI: 10.1073/pnas.1012551108.
- PIERREHUMBERT, JANET B. 2001. Exemplar dynamics: Word frequency, lenition and contrast. In Bybee & Hopper, 137–57.

- PIERREHUMBERT, JANET B. 2002. Word-specific phonetics. *Laboratory phonology 7*, ed. by Carlos Gussenhoven and Natasha Warner, 101–40. Berlin: De Gruyter. DOI: 10.1515/9783110197105.1.101.
- PIERREHUMBERT, JANET B. 2006. The next toolkit. *Journal of Phonetics* 34.516–30. DOI: 10.1016/j.wocn.2006.06.003.
- PITT, MARK A.; KEITH JOHNSON; ELIZABETH HUME; SCOTT KIESLING; and WILLIAM RAYMOND. 2005. The Buckeye corpus of conversational speech: Labeling conventions and a test of transcriber reliability. *Speech Communication* 45.89–95. DOI: 10.1016/j.specom.2004.09.001.
- POSTAL, PAUL M. 1966. On so-called ‘pronouns’ in English. *17th Annual Round Table (Monograph series on language and linguistics)*, ed. by Francis P. Dineen, 177–206. Washington, DC: Georgetown University Press. Online: https://repository.library.georgetown.edu/bitstream/handle/10822/555459/GURT_1966.pdf.
- R CORE TEAM. 2019. R: A language and environment for statistical computing. Vienna: R Foundation for Statistical Computing. Online: <https://www.R-project.org/>.
- RICKFORD, JOHN R.; ARNETHA BALL; RENEE BLAKE; RAINA JACKSON; and NOMI MARTIN. 1991. Rappin on the copula coffin: Theoretical and methodological issues in the analysis of copula variation in African-American Vernacular English. *Language Variation and Change* 3.103–32. DOI: 10.1017/S0954394500000466.
- SADLER, LOUISA. 1998. English auxiliaries as tense inflections. *Essex Research Reports in Linguistics* 24.1–15.
- SCHIEBMAN, JOANNE. 2000. *I dunno*: A usage-based account of the phonological reduction of *don't* in American English conversation. *Journal of Pragmatics* 32.105–24. DOI: 10.1016/S0378-2166(99)00032-6.
- SCHÜTZE, HINRICH. 1995. Distributional part-of-speech tagging. *EACL '95: Proceedings of the seventh conference of the European Chapter of the Association for Computational Linguistics*, 141–48. DOI: 10.3115/976973.976994.
- SCHÜTZE, HINRICH; MICHAEL WALSH; BERND MÖBIUS; and TRAVIS WADE. 2007. Towards a unified exemplar-theoretic model of phonetic and syntactic phenomena. *Proceedings of the 29th annual meeting of the Cognitive Science Society (CogSci 2007)*, 1461–66.
- SELKIRK, ELISABETH O. 1972. *The phrase phonology of English and French*. Cambridge, MA: MIT dissertation. Online: <http://hdl.handle.net/1721.1/14788>.
- SELKIRK, ELISABETH O. 1984. *Phonology and syntax*. Cambridge, MA: MIT Press.
- SELKIRK, ELISABETH O. 1996. The prosodic structure of function words. *Signal to syntax: Bootstrapping from speech to grammar in early acquisition*, ed. by James L. Morgan and Katherine Demuth, 187–214. Mahwah, NJ: Lawrence Erlbaum.
- SELKIRK, ELISABETH O. 2011. The syntax-phonology interface. *The handbook of phonological theory*, 2nd edn., ed. by John Goldsmith, Jason Riggle, and Alan Yu, 435–83. Oxford: Blackwell.
- SELLS, PETER. 1993. Juncture and the phonology of auxiliary reduction in English. *University of Massachusetts Occasional Papers in Linguistics* 8.76–105. Online: <https://scholarworks.umass.edu/umop/vol8/iss1/6>.
- SEYFARTH, SCOTT. 2014. Word informativity influences acoustic duration: Effects of contextual predictability on lexical representation. *Cognition* 133.140–55. DOI: 10.1016/j.cognition.2014.06.013.
- SHANNON, C. E. 1948. A mathematical theory of communication. *Bell System Technical Journal* 27.623–56. DOI: 10.1002/j.1538-7305.1948.tb01338.x.
- SHAO, ZESHU; JEROEN VAN PARIDON; FENNA POLETIEK; and ANTJE S. MEYER. 2019. Effects of phrase and word frequencies in noun phrase production. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 45.147–65. DOI: 10.1037/xlm0000570.
- SIMONS, GARY F., and CHARLES D. FENNIG. 2018. *Ethnologue: Languages of the world*. 21st edn. Dallas: SIL International. Online: <http://www.ethnologue.com>.
- SÓSKUTHY, MÁRTON, and JENNIFER HAY. 2017. Changing word usage predicts changing word durations in New Zealand English. *Cognition* 166.298–313. DOI: 10.1016/j.cognition.2017.05.032.
- SPARKS, RANDALL B. 1984. Here’s a few more facts. *Linguistic Inquiry* 15.179–83. Online: <https://www.jstor.org/stable/4178377>.
- SPENCER, ANDREW. 1991. *Morphological theory: An introduction to word structure in generative grammar*. Oxford: Basil Blackwell.

- SPENCER, JESSICA. 2014. *Stochastic effects in the grammar: Toward a usage-based model of copula contraction*. Stanford, CA: Stanford University dissertation.
- STERNBERG, SAUL; RONALD L. KNOLL; STEPHEN MONSELL; and CHARLES E. WRIGHT. 1988. Motor programs and hierarchical organization in the control of rapid speech. *Phonetica* 45.175–97. DOI: 10.1159/000261825.
- STERNBERG, SAUL; STEPHEN MONSELL; RONALD L. KNOLL; and CHARLES E. WRIGHT. 1978. The latency and duration of rapid movement sequences: Comparisons of speech and typewriting. *Information processing in motor control and learning*, ed. by George E. Stelmach, 117–52. New York: Academic Press.
- SWEET, HENRY. 1890. *A primer of spoken English*. Oxford: Clarendon.
- SZMRECSÁNYI, BENEDIKT. 2004. On operationalizing syntactic complexity. *Le poids des mots: Proceedings of the 7th International Conference on Textual Data Statistical Analysis, Louvain-la-Neuve*, vol. 2, 1032–39.
- SZMRECSÁNYI, BENEDIKT. 2005. Language users as creatures of habit: A corpus-based analysis of persistence in spoken English. *Corpus Linguistics and Linguistic Theory* 1.113–50. DOI: 10.1515/cllt.2005.1.1.113.
- TODD, SIMON. 2019. *The listener in language change*. Stanford, CA: Stanford University dissertation.
- TODD, SIMON; JANET B. PIERREHUMBERT; and JENNIFER HAY. 2019. Word frequency effects in sound change as a consequence of perceptual asymmetries: An exemplar-based model. *Cognition* 185.1–20. DOI: 10.1016/j.cognition.2019.01.004.
- TREMBLAY, ANTOINE, and R. HARALD BAAYEN. 2010. Holistic processing of regular four-word sequences: A behavioral and ERP study of the effects of structure, frequency, and probability on immediate free recall. *Perspectives on formulaic language: Acquisition and communication*, ed. by David Wood, 151–73. New York: Continuum.
- WALSH, MICHAEL; BERND MÖBIUS; TRAVIS WADE; and HINRICH SCHÜTZE. 2010. Multi-level exemplar theory. *Cognitive Science* 34.537–82. DOI: 10.1111/j.1551-6709.2010.01099.x.
- WEDEKIND, JÜRGEN, and RONALD M. KAPLAN. 2012. LFG generation by grammar specialization. *Computational Linguistics* 38.867–915. DOI: 10.1162/COLI_a_00113.
- WESCOAT, MICHAEL T. 2002. *On lexical sharing*. Stanford, CA: Stanford University dissertation.
- WESCOAT, MICHAEL T. 2005. English nonsyllabic auxiliary contractions: An analysis in LFG with lexical sharing. *Proceedings of the LFG '05 Conference*, 468–86. Online: <http://web.stanford.edu/group/cslipublications/cslipublicationsLFG/10/pdfs/lfg05wescoat.pdf>.
- WESCOAT, MICHAEL T. 2007. Preposition-determiner contractions: An analysis in optimality-theoretic lexical-functional grammar with lexical sharing. *Proceedings of the LFG '07 Conference*, 439–59. Online: <http://web.stanford.edu/group/cslipublications/cslipublicationsLFG/12/papers/lfg07wescoat.pdf>.
- WESCOAT, MICHAEL T. 2009. Udi person markers and lexical integrity. *Proceedings of the LFG '09 Conference*, 604–22. Online: <http://web.stanford.edu/group/cslipublications/cslipublicationsLFG/14/papers/lfg09wescoat.pdf>.
- WILDER, CHRIS. 1997. English finite auxiliaries in syntax and phonology. *Clitics, pronouns, and movement*, ed. by James R. Black and Virginia Motapanyane, 321–62. Amsterdam: John Benjamins.
- WRIGHT, SAUNDRA K.; JENNIFER HAY; and TESSA BENT. 2005. Ladies first? Phonology, frequency, and the naming conspiracy. *Linguistics* 43.531–61. DOI: 10.1515/ling.2005.43.531.
- YAGUCHI, MICHIKO. 2010. The historical development of the phrase *there's*: An analysis of the *Oxford English Dictionary* data. *English Studies* 91.203–24. DOI: 10.1080/00138380903355007.
- ZHAN, MEILIN, and ROGER LEVY. 2018. Comparing theories of speaker choice using a model of classifier production in Mandarin Chinese. *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, vol. 1 (long papers)*, 1997–2005. DOI: 10.18653/v1/N18-1181.
- ZHAN, MEILIN, and ROGER LEVY. 2019. Availability-based production predicts speakers' real-time choices of Mandarin classifiers. *Proceedings of the 41st annual meeting of the*

Cognitive Science Society (CogSci '19), 1268–74. Online: <https://cogsci.mindmodeling.org/2019/papers/0231/0231.pdf>.

ZWICKY, ARNOLD M. 1970. Auxiliary reduction in English. *Linguistic Inquiry* 1.323–36. Online: <https://www.jstor.org/stable/4177570>.

ZWICKY, ARNOLD M. 1977. On clitics. Columbus: The Ohio State University, ms. [Reproduced by the Indiana University Linguistics Club, 1977.]

ZWICKY, ARNOLD M., and GEOFFREY K. PULLUM. 1983. Cliticization vs. inflection: English *n't*. *Language* 59.502–13. DOI: 10.2307/413900.

Department of Linguistics and The Center for
the Study of Language & Information (CSLI)
Stanford University
Stanford, CA 94305
[bresnan@stanford.edu]

[Received 23 October 2018;
revision invited 24 April 2019;
revision received 6 November 2019;
revision invited 11 April 2020;
revision received 15 June 2020;
revision invited 15 September 2020;
revision received 22 September 2020;
accepted 29 September 2020]