

Frequency and variation in English subject-verb contraction*

Joan Bresnan & Jessica Spencer
Stanford University

November 23, 2013

Frequency effects in syntax are a focus of intense interest in language and cognition (Bybee & Thompson 1997; Bybee & Sheibman 1999; Bybee & Hopper, eds.; 2001, Bybee & McClelland 2005; Gahl & Yu, eds., 2006; Diessel 2007; Arnon & Snider 2010; Wiechmann et al., eds., 2013). Lexical frequency effects that distinguish ‘stored’ from ‘computed’ structures (e.g. Pinker & Ullman 2002) are compatible with many generative models of language representation and processing, while syntactic frequency effects, by which we mean those that occur with multiword sequences, are compatible with many experience-based models of linguistic knowledge (e.g. Arnon & Snider 2010). We would not expect to find syntactic frequency effects if, in the words of Diessel (2007: 108), syntax forms “a closed and stable system that is not affected by pragmatic and psycholinguistic principles involved in language use.” But we would expect syntactic frequency effects if, in contrast, syntactic structures “emerge from processing linguistic data” as part of “a dynamical system that is constantly changing by virtue of psychological processes involved in language use.”

Many studies of frequency effects on ambiguous syntactic structures exist (see Diessel 2007 for a review), but in these, different syntactic structures covary with different semantics. Variation is a crucial domain for investigating syntactic frequency effects, because it focuses on syntactic differences isolated from covarying semantics and lexical instantiations. A central question, then, is whether frequency effects occur in syntactic variation.

*We are grateful to Sam Bowman and Arto Anttila for their substantive contributions to the data collection, annotation, and conceptualization of this study. This article is based on work supported by National Science Foundation Grant No. BCS-1025602.

A number of previous studies have reported frequency effects on syntactic variation. Cases include the choice between subjunctive and indicative complements in Canadian French (Poplack 1989), the reduction of *don't* in English (Bybee & Sheibman 1999), vowel centralization in dative *give* constructions in New Zealand English (Hay & Bresnan 2006), variation between future expressions in English such as *will*, *be going to* (Torres Cacoullos & Walker 2009), subject-pronoun omission in Spanish (Erker & Guy 2012, but cf. Bayley, Greer, & Holland 2012), and subject-verb contraction in English (Bybee & Thompson 2000; Krug 1998, 2003; Frank & Jaeger 2008; Barth 2011; MacKenzie 2012). Yet the question remains: To what extent do these phenomena reflect “living” frequency effects on productive syntactic variation? “Living” frequency effects occur during compositional language processing and may be contrasted with the products of diachronic coalescence of several forms into fused lexical units (e.g. Bybee & Sheibmann 1999, Bybee 2001) or unproductive stored routines (Bybee & Thompson 2000: 384).

The present study investigates whether there are living syntactic frequency effects in English subject-verb contraction, a well-studied area of grammar prone to variation. Several examples collected from the Buckeye corpus of English (Pitt et al. 2007) illustrate the phenomenon in (1a)–(1d):

- (1) a. my best friend **is** a realtor uh **she's** ...
 b. that **is** uh <VOCNOISE> **that's** a prejudice thing
 c. this girl **is** almost creepy **she's** so ...
 d. they **have** bought generators **they've** bought water

Note that the same speaker may vary contraction within the same utterance with the same host and verb. (The term ‘host’ designates the preceding word with which the verb contracts.)

1 Previous findings from studies of contraction

Subject-verb contraction is restricted to the highest-frequency verbs in English: the copula, auxiliaries, and modal verbs. These verbs retain syntactic properties such as inversion and *not* placement which were prevalent in Middle English, reflecting “the conserving effect” of high frequency forms in syntax (Bybee & Thompson 2000).

Previous corpus studies have found a number of predictors of variable subject-verb contraction, briefly summarized here.

Preceding segment phonology: The segmental features of the host (e.g. *they* in *they've*) play a large role; those ending in vowels favor contraction (Labov 1969, MacKenzie 2012).

Following constituent category: Copula contraction varies given the grammatical category of the complement constituent, more often with verbal complements than with nominal (Labov 1969, Rickford et al. 1991, McElhinny 1993, MacKenzie 2012; also Barth 2011).

Length of subject in words: Length in words of the subject of *is* is inversely related to subject-verb contraction (Frank & Jaeger 2008, Barth 2011, MacKenzie 2012).

Information load: Three *n*-gram measures of information load are inversely related to contraction of host and verb: (Frank & Jaeger 2008). These respectively denote the information load on the verb in the context of the next word, in the context of the host and word previous to the host, and the information load on the next word in the context of the host and verb. In all cases, higher information load on the verb or following word is associated with a lower probability of contraction. Because information load is defined in terms of (the binary logarithm of) the inverse of the probability, information load decreases as probability increases, capturing the intuition that highly probable, predictable, or redundant events are relatively uninformative, and thereby subject to greater reduction.

Frequency: Higher joint probability of host and verb are directly related to subject-verb contraction in both spoken and written English corpora (Krug 1998, 2003; Barth 2011). Higher probability of the host is also directly related to subject-verb contraction in a spoken English corpus (Frank & Jaeger 2008). High joint probability of BE and the following word (particularly *gonna*) increases contraction (Labov 1969, McElhinney 1993, MacKenzie 2012).

The frequency effect on subject-verb contraction is contested. Krug (1998) was the first to observe that phonological explanations fail to account for the variability in contraction of *have* among phonologically similar hosts and that the 'string frequency' (joint probability) of the host and verb accounts for the variation that

phonology alone cannot in spoken mainstream British English.¹ The phonologically similar contexts in question are all personal pronouns ending in vowels before contracted 've, but he extends his study to other subject-verb combinations, admitting the “distorting effects” of other uncontrolled variables in these cases. Barth (2011) confirms effects of string frequency (pronoun + verb-form) on English subject-verb contraction in both historical written data and in transcriptions of contemporary television and radio news programs from the Corpus of Historical American English and the Corpus of Contemporary American English (Davies 2008–, 2010–). But MacKenzie (2012) fails to replicate Krug’s findings of a string frequency effect, in a large sample from the Penn Treebank Switchboard corpus (Marcus, Marcinkiewicz, and Santorini 1993) of subject pronoun-verb combinations that undergo contraction. She suggests that Krug’s string frequency is only a binary factor that distinguishes the highest cases on the joint probability scale (pronouns) from the lowest (the lexical subjects).²

In all of these previous studies, the reported frequency effects are largely driven by pronoun subjects. But some researchers of formal theories of grammar have independently proposed that contracted pronouns and auxiliaries in English may be morphologically fused (Spencer 1991) or incorporated into a single word, whether by verb incorporation into the subject (Sadler 1998), lexical sharing by verb and subject pronoun preterminals (Wescoat 2005), or by incorporation of the subject pronoun into the head verb (Bender & Sag 2001). On these theories, the apparent effects of item frequencies on a syntactic alternation could reduce to the effects of lexical frequency of the stored, coalesced subject + verb forms. This theoretical possibility raises again the central question: *Do host + verb frequencies affect subject-verb contractions in living multi-word syntactic alternations in spontaneous uses of language?* This is the fundamental question the present study addresses.

¹—drawn from the London Lund Corpus of Spoken English (Svartik 1959–) and the spoken component of the Bank of English Corpus, jointly owned by HarperCollins Publishers and the University of Birmingham.

²Unlike Barth (2011) MacKenzie appears not to have normalized the host frequencies by the frequencies of the specific verb forms they occur with (e.g. *are, is am* vs. *be*). Further, unlike both Frank & Jaeger (2008) and Barth (2011), MacKenzie (2012) does not include her frequency measures as a predictor in her models of contraction.

2 Data

To investigate syntactic frequency effects in subject-verb contraction, we used the Buckeye Corpus of English (Pitt et al. 2007), first to see if we could replicate Krug’s findings with HAVE and BE in conversational Midwestern North American English, and then to study lexical (non-pronoun) hosts to look for living effects of frequency information on contraction throughout the lexicon.

The Buckeye Corpus consists of interviews with 40 people, constituting 300,000 words in all. The speakers are Caucasian, long-time local residents of Columbus, Ohio. The language is unmonitored casual speech. The corpus is stratified by age and gender: 20 old, 20 young, 20 male, 20 female. The corpus provides high-quality phonetic labeling, part-of-speech (POS) tagging, and alignment of words and phones aligned with sound waves.

2.1 Defining the contraction variable

Subject verb contraction involves the deletion of the onset (if one exists) and nucleus of English auxiliaries BE, HAVE, *will*, *shall*, and *would*. In the present study only the asyllabic forms of the auxiliaries *have* and *has* were considered to be contracted. Table 1 shows the Buckeye phonetic transcriptions that were used to define our dependent variables.

Table 1: Full and contracted forms considered in Experiments 1 & 2

Transcribed	Full form	Contracted form
am	[æm], [aɪm], [ɛm], [ɪm]	[m]
are	[ɹ], [ɹr] [ɑr], [əɹ], [ɛ] [ɛr] [əʊr] [ʊr]	[r]
is	[ʌs], [ʌz], [ɛs], [ɛz], [ɪs], [ɪz], [iz], [ʊs], [ʊz]	[s] or [z]
has	[hæz], [hɛs], [hʌz] [hɪs]	[s] or [z]
have	[hæv]	[v]

2.2 Collection method

Using a perl script, we collected from the Buckeye corpus all examples of full and contracted present tense BE and HAVE forms by extracting all occurrences of *am*, *are*, *is*, and words containing *'m*, *'re* or *'s* along with four context items on either side. We

read through all 8,690 cases originally extracted of *'s/is* forms and manually excluded instances of possessive *'s* and *let's*, and separated contracted forms of *is* and *has*, both transcribed *'s*, making use of POS tags. Of these, 6,282 were cases where *is/'s* variation is possible and 343 were instances of contracted *has* (with 87 instances of full *has* being extracted separately).

2.3 Exclusion criteria

Data selection followed MacKenzie (2012). Hosts were limited to subject NP hosts, as non-subject NP hosts block or do not exhibit enough variation. This excludes sentential, as well as preposed PP, adverbials, and VP hosts. Additionally focus-like constructions that begin with *the thing is...* were excluded as they also prohibit contraction. Auxiliaries immediately preceding a gap were also excluded. Negated auxiliaries were not included, as the envelope of variation differs in these cases (i.e. *he's not* vs. *he isn't* (Tagliamonte & Smith 2002, Yaeger-Dror et al. 2002). Finally, all instances of *has* or *is* following a sibilant were excluded, as the asyllabic forms do not occur in that environment in our data.

Phonetic transcription was used to ensure that occurrences of BE forms that were lexically transcribed as contracted were in fact also phonetically contracted variants. If a reduced form of the vowel was phonetically transcribed for an orthographically transcribed token (e.g. [deyvəz] vs. [deyvz] for *Dave's*), it was excluded from analysis. (The numbers excluded for this reason: *I'm* = 7, *we're* = 2, *they're* = 4, *'s (is)* = 2.) Conversely, auxiliaries orthographically transcribed as full but phonetically transcribed as contracted were also excluded (*am* = 1, *are* = 4, *is* = 54, *has* = 3). In the case of *has* and *have*, all 'intermediate' forms (MacKenzie 2012), those only missing an onset, were excluded (*has* forms = 32, *have* forms = 1).

Other reasons for exclusion involve inability to rigorously define subject or complement categories within the given nine word context window, as well as false starts or repetitions. This left 5,507 full or contracted *is*, 430 *has*, 1,128 *am*, 2,848 *are*, and 1,773 *have*.

3 Experiment 1: Contraction with pronoun subjects

Given Krug's (1998) results, the proportion of HAVE contractions should be directly related to the frequency of co-occurrence between *host* and verb. However, the relation between these statistics is highly nonlinear. We therefore transformed the proportion of contracted forms to the log odds of contraction, and the string frequency

of the host + verb to the (log of) the string frequency normalized by the specific verb form, *has*/'s or *have*/'ve in this case:

$$\frac{\text{host} + \text{VERB form}}{\text{VERB form}}$$

This transformation is the log conditional probability of the host given the verb form (cf. Jurafsky et al. 2000). It provides a very good fit to the contraction data for HAVE, as shown by the lowess smoother line indicating the trend of the data points in the left panel of Figure 1. The right panel of Figure 1 shows the results for BE contraction with frequent pronoun hosts in the Buckeye corpus, using the same transformations.³

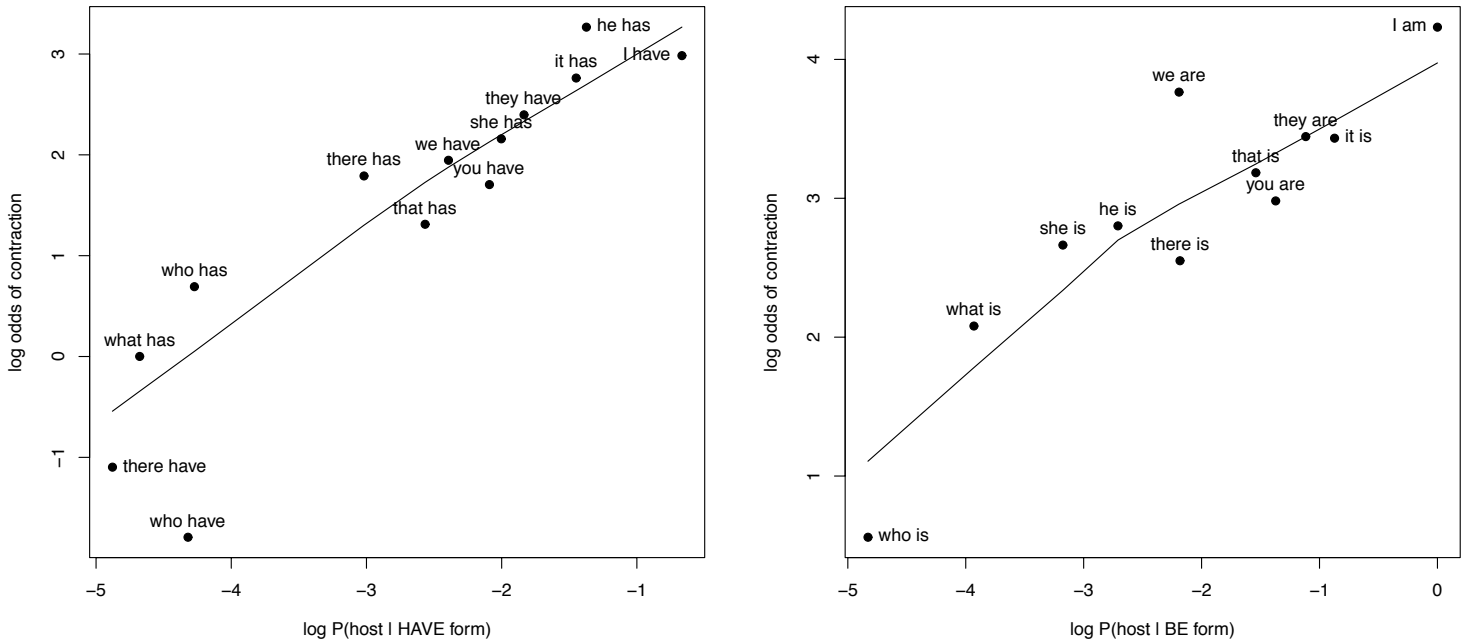


Figure 1: Change in probability of contraction as a function of change in conditional probability of pronoun host given verb form

³By simple linear regressions the log odds of contraction of HAVE = $4.1865 + 1.0211 \times \log P(\text{pronoun}|\text{HAVE form})$, $F(1, 11) = 54.24$, $p < 0.0001$, $R^2 = 0.8161$, and the log odds of contraction of BE = $4.2228 + 0.6170 \times \log P(\text{pronoun}|\text{BE form})$, $F(1, 9) = 35.91$, $p < 0.001$, $R^2 = 0.7773$.

3.1 Discussion of Experiment 1

These results confirm Krug’s earlier findings, and improve upon them by considering the differences in the relative frequencies of the different BE and HAVE forms. But they are open to question on at least two fronts. First, pronoun subjects are highly frequent hosts of contraction and are liable to coalesce with the highly frequent auxiliary verbs they occur with, as argued in the previous literature reviewed in section 1. Hence the frequency effects shown in Figure 1 could ultimately reduce to lexical frequency effects for stored pronoun + verb combinations. Subject-verb contraction would then not offer a case of frequency effects on productive *syntactic* variation.

Second, independently of the pronoun-auxiliary coalescence theories, string frequency could for whatever reason merely affect contraction with the highest frequency subjects, which are pronouns. If so, it might have no effect on the lowest frequency subjects, which are lexical. This could be viewed as a version of MacKenzie’s (2012) proposal that string frequency separates the most frequent subject-verb sequences from the least frequent in a binary way.

These considerations led us to our second experiment.

4 Experiment 2: Contraction with lexical subjects

To probe for “living” syntactic frequency effects on subject-verb contraction, we decided to investigate contraction with lexical subjects of the verb *is*, the only verb that contracts with a wide variety of hosts sufficiently often in our dataset to permit reliable analysis. After we excluded all personal pronouns, demonstrative pronouns, interrogative (*wh*-) pronouns, and expletive pronoun subjects of *is*’s, 629 instances of variable *is*’s remained, of which 283 were contracted.

Several examples from the dataset are given in (2a)–(2e):

- (2) a. yeah my **dad’s** a chauvinist pig
- b. <SIL> yknow <SIL><VOCNOISE> **life’s** a struggle <SIL>
- c. the **powder cocaine’s** another <VOCNOISE> you know
- d. my brother in law **is** gay
- e. my other **brother in law’s** Arab

The variety of hosts in the lexical subject dataset is substantial: there are 314 different lexical hosts (types). However, 212 of these are single instances, the least

frequent in our Buckeye sample. If the single instances vary randomly in contracting with *is*, it would be a drag on the predictability of contraction by any model. In fact, though, 57 occur with contracted 's and 155 occur with the full form. This is a lower rate of contraction than occurs with hosts having two or more instances; the latter have 226 contracted and 191 full. These summary statistics suggest that there is a robust frequency effect on contraction in the lexical host *is* data, despite the large number of singleton and very infrequent hosts, which add noise to the data.

4.1 Predictors

We coded the data for known predictors of contraction (section 1), operationalized as follows.

- **Preceding segment phonology:** We coded for whether the host ended phonologically in a consonant (TRUE) or not (FALSE).
- **Length of subject in words:** We manually identified the subject constituent preceding *is* or 's and coded it for its length in orthographic words: 1, 2, 3, or 4 or more (4+).
- **Following constituent category:** We manually identified the following constituent category of the complement of *is*/'s and categorized it according to the three-way collapsed classification argued for by MacKenzie (2012: 167): nominal, verbal, and adjectival.
- **Is type:** We also coded a simpler binary distinction between *is*/'s as an auxiliary or as a copula which implicitly classifies the complement constituent category as verbal (including *gonna*) or non-verbal (Rickford et al. 1991).
- **Information load:** Following Frank & Jaeger (2008), we collected the unigram, bigram, and trigram statistics of the Buckeye corpus. We then calculated and transformed the conditional probabilities needed to define the three information-load predictors which they found to be significantly related to contraction in the SWITCHBOARD CORPUS: $I(verb|next)$, $I(verb|prev, host)$ and $I(next|host, verb)$.
- **Frequency:** We calculated the joint probabilities of the host and *is*/'s and of *is*/'s and the following word, respectively $P(host, verb)$ and $P(verb, next)$.

- **Speech rate:** Speech rate is a predictor of phonetic reduction but not necessarily of contraction (cf. Kaisse 1985, Frank & Jaeger 2008, MacKenzie 2012). We coded it as a control, using a perl script developed by Bowman (2012) to compute local speech rate as the nuclei per second averaged within a ten-word window of *is*/'s and normalized by the overall average rate for each speaker.

4.2 Model selection

We modeled the data using multiple logistic regression treating the binary outcome variable, Contracted or not, as a linear combination of all of our predictors. We defined the contracted form 's as 1 and the full form as 0 in the model.⁴

To satisfy model assumptions, we logged, centered, and standardized the numerical predictors to reduce skewing. We also centered the binary factors. To reduce multicollinearity, we decorrelated two pairs of the transformed numerical predictors: $I(next|host, verb)$ from $P(host, verb)$, and $P(verb, next)$ from $P(host, verb)$. To eliminate collinearity between the overlapping factors of following constituent category and *is*-type, we decided to drop one.⁵ Because these alternative predictors are essentially equivalent, we decided to drop *is*-type for the sake of comparability with MacKenzie (2012).

From our working model containing the predictors above, we eliminated predictors whose slope estimates were indistinguishable from noise ($\beta < s.e.$). That included the previous segment phonology and speech rate predictors.

Finally, we corrected for data dependencies from the same speakers by bootstrapping ($B=200$) with cluster replacement from our working model (Harrell 2001). The Wald statistics for this corrected model are shown in Table 2.⁶

We then subjected the corrected model to bootstrap validation ($B = 1000$) to test for possible overfitting. The validated model has good accuracy ($C = 0.8003$,

⁴Note this is the opposite of Frank & Jaeger's (2008) model, which included pronoun hosts as well as multiple verb forms that vary with host features.

⁵To choose which to drop, we used likelihood ratio tests to compare how much each contributed to a full model of the data containing all other predictors. In this comparison, neither contributed significantly to goodness of fit in a model including the other, though the *is*-type factor was slightly better: $L.R. \chi^2(2) = 2.979, Pr(> \chi^2) = 0.08627$ for the model dropping the following constituent category factor vs. $L.R. \chi^2(1) = 1.943, Pr(> \chi^2) = 0.22548$ for the model dropping the *is*-type factor. A comparison of AICc values for the two *non-nested* models identical except for the choice of following constituent category or *is*-type showed the same: *Is*-type produced a better (lower) AICc value: 685.1922 vs. 687.221.

⁶"ctr", "std", "res" respectively denote "centered", "standardized", and "residualized".

Table 2: Model parameter estimates for lexical subject contraction with *is*

	Coefficient	Standard Error	Wald Z	Pr(> Z)
Intercept	−0.7275	0.1990	−3.66	0.0003
Following = adjectival	0.2628	0.2060	1.28	0.2020
Following = verbal	1.3211	0.2788	4.74	<0.0001
Subject length (ctr)	−0.5975	0.1260	−4.74	<0.0001
$\log P(host, verb)$ (std)	0.5970	0.1362	4.38	<0.0001
$\log P(verb, next)$ (std,res)	0.7705	0.2073	3.72	0.0002
$I(verb prev, host)$ (std)	0.1521	0.1149	1.32	0.1858
$I(verb next)$ (std)	−0.2714	0.1445	−1.88	0.0603
$I(next host, verb)$ (std,res)	−0.6316	0.2095	−3.01	0.0026

optimism 0.0214) and accounts for about a third of the variance ($R^2 = 0.3382$, optimism 0.0261). The R^2 value suggests that the data are very noisy, as we have already noted.

The significant partial effects of the final model are plotted with 95% confidence intervals in Figure 2.

4.3 Interpretation of the model

The top left panel in Figure 2 shows that a following nominal complement to *is* is associated with lower probability of contraction with a lexical subject than a verbal complement is (cf. Labov 1969, Rickford et al 1991, McElhinney 1993, MacKenzie 2012). The 0 level on the y -axis signifies even odds of contraction; the odds of contraction with verbal complements are above the 0 level, while the odds of contraction with the other complement categories are below it. The vertical lines through the three points represent the ranges within the values of the parameters lie with 95% probability. So, at least for lexical subjects and for the three-way categorization of the following complement category we used, adjectival complements do not differ significantly from the nominal complements in their relation to contraction of *is* with lexical hosts.

The top right panel in Figure 2 shows that subject word count is inversely related to contraction of *is* (Frank & Jaeger 2008, Barth 2011, MacKenzie 2012). Because we centered the Subject word count predictor, the average word count is at 0 on the x -axis. Note that the average is associated with lower than even odds of contraction of *is* (below 0 on the y -axis). As the length value increases above average, the odds

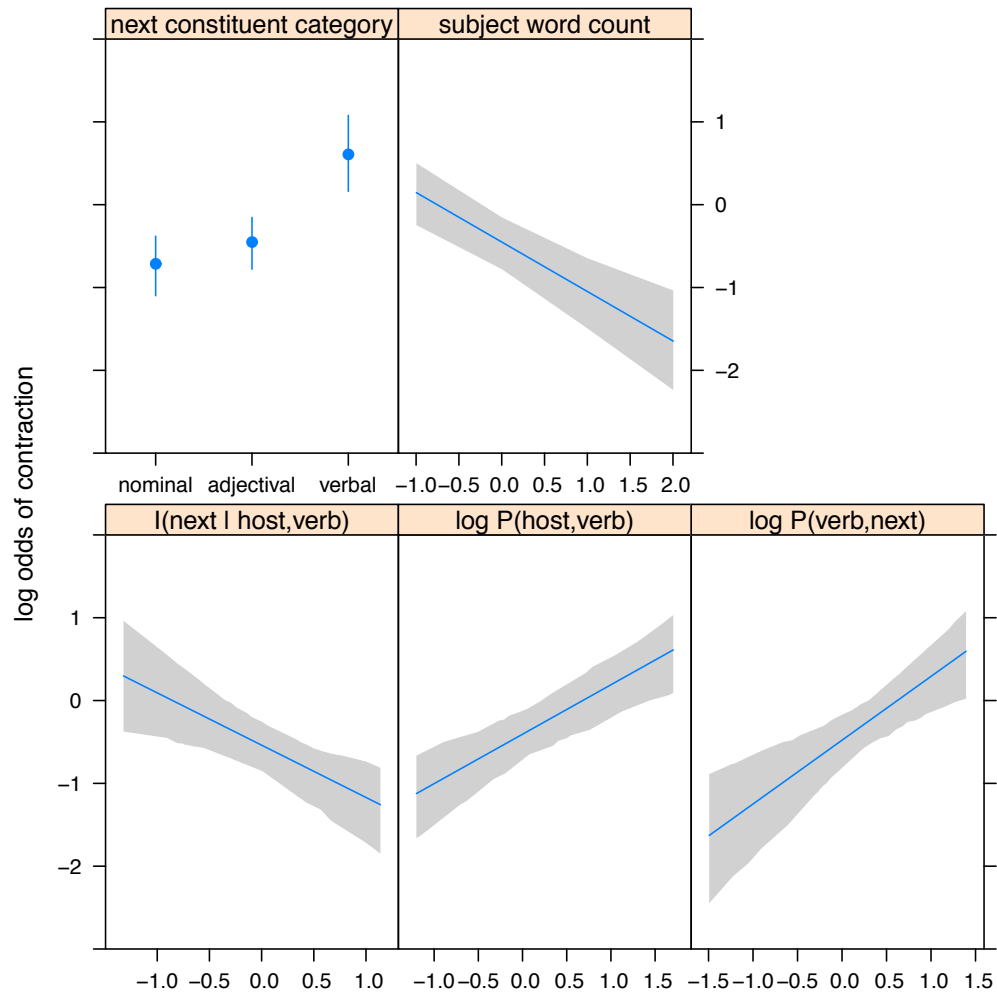


Figure 2: Partial effects of model of lexical subject contraction with *is*

of contraction decline further; likewise, as the length decreases below the average, the odds of contraction increase.

In each of the lower panels, the x -axis predictor has been standardized by dividing the measure by its standard deviation, so that the three lower panels are all on the same scale. In all three lower panels both x -axis and y -axis are log scales. The lines therefore show the proportion of change in y as a function of change in x .

The lower left panel shows that the greater the information load on the upcoming word in the context of the host and *is*, the less likely it is for contraction to occur (Frank & Jaeger 2008). Recall that the information load is defined as the reciprocal of the probability, so increasing information load implies decreasing probability. Therefore the less probable the upcoming word is following the host and verb, the more likely speakers are to use the full, uncontracted form *is*.

The middle lower panel shows that the joint probability of host and *is*/'s is directly related to the contraction rate (Krug 1998, Bybee 2001, Barth 2011). This is the “string frequency” effect: as the joint probability of host and verb increase, contraction increases.

Finally, the lower right panel shows that the joint probability of *is*/'s and the following word is directly related to contraction rate, even when the syntactic category of the following constituent is controlled for. We included this predictor to capture the effects of particular items such as *'s gonna* and *'s like*, which have been studied by sociolinguists (Labov 1972, Rickford et al 1991, McElhinny 1993).

4.4 Discussion of Experiment 2

Experiment 2 shows that joint probability of host and BE form affects contraction over and above the effects of word-level information load, syntax, and phonology found in previous studies. Moreover, all of the significant effects we found are similar in direction to those occurring in previous datasets dominated by pronoun hosts.

Four predictors dropped out of our model that were found to be reliable in previous studies: speech rate, preceding segment phonology, and information load on the verb in the context of the next word [$I(verb|next)$], and in the context of the host and word previous to the host [$I(verb|prev, host)$]. It is possible that the small size of our dataset is insufficient for detecting these effects. This is likely in the case of the predictor $I(verb|prev, host)$. The estimated slope for this predictor is in the theoretically expected direction. It probably fails to reach significance at the 95% threshold because the effect is too small to detect reliably in our dataset; it would possibly play a role in a larger set of data. However, the same cannot be said for the other information load predictor $I(verb|prev, host)$ that dropped out.

As for speech rate, Frank & Jaeger (2008) similarly found an absence of the effect they expected on their view that the phonetic reduction in fast speech is subject to the same principle of Uniform Information Density as the choice of reduced allomorphs in subject-verb contraction. Yet phonological studies of subject-verb contraction have distinguished between fast speech phenomena, such as palatalization and flapping in English, which are sensitive to syllable structure and apply both within and across word boundaries, and syntactically constrained processes of cliticization and subject-verb contraction (Kaisse 1985).

It is possible that the effects in previous studies that dropped out of the present study were driven by the high-frequency pronouns, which constitute a small finite inventory of types. For example, a preceding consonant does not inhibit contraction in the dataset of the present study (cf. Labov 1996). Perhaps without the higher frequency pronominal subjects which carried the effect before, the effect disappears. Most of the high-frequency pronoun subjects are open monosyllables.

5 General discussion

Our findings can be viewed as another instance of a much broader generalization that applies to phonetic reduction in probable syntactic constructions (e.g. Gahl & Garnsey 2004, Kuperman & Bresnan 2012) as well as the use of shorter morpholexical variants such as clitics and the omission of optional elements such as complementizers or relative pronouns (Jaeger 2010). In speaking, more probable units are more reduced, or in the often-cited words of Du Bois (1985: 362–363): *Grammars code best what speakers do most*—that is, grammars provide the most economical expressions for the speech functions that speakers utilize most often. Why should this be?

Our study of contraction in Buckeye provides support for several theories of this reduction phenomenon. The first is *Routinization*. Production of frequent collocations leads to routinization of motor plans for articulation, which eventually become grammaticalized (Bybee & Thompson 2000, Bybee 2001). Our high-frequency pronoun studies in Experiment 1 enhance the support for this theory.

The second is *Availability*. On this theory speakers use longer (unreduced) expressions to “buy time” to prepare for upcoming infrequent or difficult continuations (Ferreira & Dell 2000). Infrequent expressions are cognitively less available. The significance of our next-word based predictors, joint probability of verb and next word [$\log P(\text{verb}, \text{next})$], and information load on the next word [$I(\text{next}|\text{host}, \text{verb})$] in our Experiment 2 enhances support for this theory, as also suggested in the study

by Frank & Jaeger (2008).

The third theory of reduction is *Uniform Information Density*. Speakers use the minimal effort needed to ensure maximal communication in an information-theoretic sense because more predictable words are easier to comprehend (cf. Jurafsky et al. 1998; Frank & Jaeger 2008). The information-load predictor $I(next|host, verb)$ supports this theory. While the present study cannot distinguish between this and the Availability theory, Kuperman & Bresnan (2012) provide evidence from acoustic durations during incremental production that does.

The fourth theory of reduction is provided by *Exemplar-based theories* (Bybee & Hopper, eds., 2001; Gahl & Yu, eds., 2006, Walsh et al. 2010, Wiechmann et al., eds., 2013). In the words of Barth (2011), “listeners build up memories of hypo-articulated forms of frequent words, and then in turn use these memories to produce their own speech, further entrenching the idea of a lenition-bias on frequent forms...” The importance of the joint probability of host and verb predictor $[logP(host, verb)]$ in both experiments lends support this theory.

The main conclusion we draw from the experiments of the present study is that the effect of joint probability of host and BE form on contraction is not a binary high/low effect, but appears to be a pervasive property through the frequency range. Hence, joint frequency of host and auxiliary affects not just the products of diachronic coalescence of words into fused units which may be stored rather than computed compositionally, it has effects on living multiword syntactic alternations. Our findings accord with conceptions of the plasticity of grammar as a function of experience.

References

- Arnon, Inbal, and Neal Snider. 2010. More than words: Frequency effects for multiword phrases. *Journal of Memory and Language* 62(1):67–82.
- Barth, Danielle. 2011. Modeling reduction of *is*, *am* and *are* in grammaticalized constructions. *Quantitative Issues in Theoretical Linguistics (QITL)* 4 4:13.
- Bayley, Robert, Kristen Greer, and Cory Holland. 2013. Lexical frequency and syntactic variation: A test of a linguistic hypothesis. *University of Pennsylvania Working Papers in Linguistics* 19(2):4.
- Bender, Emily M, and Ivan A Sag. 2000. Incorporating contracted auxiliaries in english. *Grammatical interfaces in HPSG* 17–32.
- Bybee, Joan. 2001. *Phonology and Language Use*. Cambridge: Cambridge University Press.

- 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(2–4):381–410.
- Bybee, Joan, and Sandra Thompson. 1997. Three frequency effects in syntax. In *Proceedings of the Annual Meeting of the Berkeley Linguistics Society: General Session and Parasession on Pragmatics and Grammatical Structure*, Vol. 23, 378–388.
- Bybee, Joan L, and Paul J Hopper (Eds.). 2001. *Frequency and the Emergence of Linguistic Structure*. John Benjamins Publishing.
- Davies, Mark. 2008–. The Corpus of Contemporary American English (COCA): 400+ million words, 1990–present. *Online at* <http://www.americanacorporus.org>.
- Davies, Mark. 2010–. The Corpus of Historical American English (COHA): 400+ million words, 1810–2009. *Online at* <http://www.americanacorporus.org>.
- Diessel, Holger. 2007. Frequency effects in language acquisition, language use, and diachronic change. *New Ideas in Psychology* 25(2):108–127.
- DuBois, Jack. 1985. Competing motivations. In J. Haiman (Ed.), *Iconicity in Syntax*, 343–365. Amsterdam: Benjamin.
- Erker, Daniel, and Gregory R Guy. 2012. The role of lexical frequency in syntactic variability: Variable subject personal pronoun expression in spanish. *Language* 88(3):526–557.
- Ferreira, Victor S, and Gary S Dell. 2000. Effect of ambiguity and lexical availability on syntactic and lexical production. *Cognitive psychology* 40(4):296–340.
- Frank, Austin, and T. Florian Jaeger. 2008. Speaking rationally: Uniform information density as an optimal strategy for language production. In *The 30th Annual Meeting of the Cognitive Science Society (CogSci08)*, 939–944, Washington, D.C., July.
- Gahl, S., and S.M. Garnsey. 2004. Knowledge of grammar, knowledge of usage: syntactic probabilities affect pronunciation variation. *Language* 80(4):748–774.
- Gahl, Susanne, and Alan C. L. Yu (Eds.). 2006. *The Linguistic Review: Special Issue on Exemplar-based Models in Linguistics*. Vol. 23.
- Harrell, Frank E. 2001. *Regression modeling strategies: with applications to linear models, logistic regression, and survival analysis*. Springer.
- Hay, Jennifer, and Joan Bresnan. 2006. Spoken syntax: The phonetics of giving a hand in New Zealand English. *The Linguistic Review* 23(3):321–349.
- Jaeger, T. Florian. 2010. Redundancy and reduction: Speakers manage information density. *Cognitive Psychology* 61(1):23–62.
- Jurafsky, Daniel, Alan Bell, Michelle Gregory, and William D. Raymond. 2000.

- Frequency and the Emergence of Linguistic Structure*. chapter Probabilistic relations between words: evidence from reduction in lexical production, 229–254.
- Kaisse, Ellen M. 1985. *Connected speech: the interaction of syntax and phonology*. Academic Press New York.
- Krug, Manfred. 1998. String frequency: A cognitive motivating factor in coalescence, language processing, and linguistic change. *Journal of English Linguistics* 26(4):286–320.
- Kuperman, Victor, and Joan Bresnan. 2012. The effects of construction probability on word durations during spontaneous incremental sentence production. *Journal of Memory and Language* 66(4):588–611.
- Labov, William. 1969. Contraction, deletion, and inherent variability of the English copula. *Language* 45(4):715–762.
- MacKenzie, Laurel E. 2012. *Locating Variation Above the Phonology*. Philadelphia: PhD. dissertation University of Pennsylvania.
- Marcus, Mitchell P, Mary Ann Marcinkiewicz, and Beatrice Santorini. 1993. Building a large annotated corpus of english: The penn treebank. *Computational linguistics* 19(2):313–330.
- McElhinny, Bonnie S. 1993. Copula and auxiliary contraction in the speech of white americans. *American Speech* 68(4):371–399.
- 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(1):89–95.
- Poplack, Shana. 1989. The inherent variability of the french subjunctive. *Theoretical Analyses in Romance linguistics* 235–263.
- 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(1):103–132.
- Sadler, Louisa. 1998. English auxiliaries as tense inflections. *Essex Research Reports in Linguistics* 24:1–15.
- Spencer, Andrew. 1991. *Morphological theory: An introduction to word structure in generative grammar*. Basil Blackwell Oxford.
- Svartvik, Jan. 1990. *The London-Lund corpus of spoken English: Description and research*. Lund University Press.
- Torres Cacoullos, Rena, and James A Walker. 2009. The present of the english future: Grammatical variation and collocations in discourse. *Language* 85(2):321–354.

- Walsh, Michael, Bernd Möbius, Travis Wade, and Hinrich Schütze. 2010. Multilevel exemplar theory. *Cognitive science* 34(4):537–582.
- Wescoat, Michael T. 2005. English nonsyllabic auxiliary contractions: An analysis in lfg with lexical sharing. In *Proceedings of the LFG05 Conference*, 468–486.
- Wiechmann, Daniel, Elma Kerz, Neal Snider, and T. Florian Jaeger (Eds.). 2013. *Language and Speech: Special Issue on Parsimony and Redundancy in Models of Language*. Vol. 56.