

Locating linguistic variation: A case study of English auxiliary contraction

Laurel MacKenzie, laurel@ling.upenn.edu

The poster is available on my website, <http://ling.upenn.edu/~laurel>. This handout clarifies and expands on the points raised there.

1. Introduction

- The phenomenon under examination in this study is the **variable contraction of English auxiliaries**:

(1) *I will ~ I'll be there.* (2) *She has ~ she's arrived.* (3) *We have ~ we've arrived.*

- Contraction is a complex process which implicates **many levels of the grammar**. Accordingly, it has been analyzed from many perspectives, from phonological (e.g. Zwicky, 1970) to morphosyntactic (e.g. Kaisse, 1983).
- However, contraction is also **variable**, a fact which has received little attention outside of the sociolinguistic literature. (A long line of sociolinguistic research, beginning with Labov, 1969, has examined the variable constraints on contraction, though it has focused almost exclusively on contraction of the copula.)
- Thus, due to its variability and to the fact that it could in principle have a number of grammatical loci, contraction is an ideal test case for questions of the **grammatical locus of linguistic variation**. To that end, this poster reports on a corpus-based study of English auxiliary contraction, examining the following:
 - The surface patterns displayed by this variable: what is the envelope of variation? At what rate(s) does contraction occur in natural speech?
 - The factors conditioning this variation: specifically, any effect of auxiliary identity and/or subject type
 - The bearings these findings have on where the locus of variation must be for this particular alternation
- I will propose that the findings presented here can be explained if we model contraction as a **two-stage process**, with variability at two levels of the grammar. Specifically, the surface data are the result of:
 1. An **allomorphic or morphophonological process** that generates full and contracted forms¹
 2. Subsequent **low-level phonological processes** that may obscure the full/contracted distinction
- These findings argue, more generally, that **careful consideration of the locus of variation is crucial for explaining surface distributions**.

2. Methodology of the Corpus Study

4898 tokens of the auxiliaries *had*, *has*, *have*, *is*, *will*, and *would* were pulled at random from the Switchboard corpus (Godfrey et al., 1992), a corpus of telephone conversations between strangers on a given topic (~240 hours/3 million words long). Tokens in the data set were uttered by 464 unique speakers. Tokens were located using transcripts, but all were hand-coded based on audio.

3. Results and Discussion

- Though previous work on auxiliary contraction has focused on a **bipartite** distinction between full and contracted, the corpus study reveals that, phonologically speaking, **three forms** are in alternation:
 1. Full: He [hæd]/[həd] been there all day. (initial consonant, audible vowel)
 2. Intermediate: He [əd] been there all day. (no initial consonant but audible vowel)
 3. Contracted: He[d] been there all day. (no initial consonant, no vowel)
- As will be discussed in Sections 3.1 and 3.2, two factors conditioning the variation between full, intermediate, and contracted emerge from the data:

Auxiliary tense: past vs. non-past

Subject type: pronoun vs. full noun phrase

¹I'm remaining agnostic about the exact nature of this higher-level process for now.

3.1. Distribution of forms after vowel-final pronoun subjects

Figure 1 shows the distribution of forms of the six auxiliaries under examination after vowel-final pronouns.

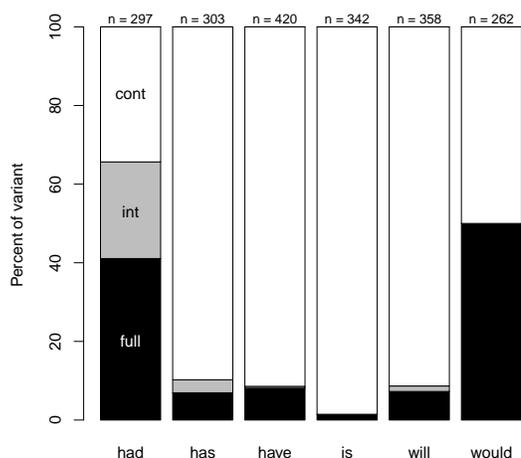


Figure 1: Distribution of forms after vowel-final pronoun subjects. Black = full, gray = intermediate, white = contracted.

- Pronouns were restricted to those that end in vowels (in other words, tokens in which an auxiliary followed *it* were omitted) in order to exclude cases in which an auxiliary was barred from close attachment to its host for phonotactic reasons (such as *it'll* [ɪrəl] and *it'd* [ɪrəd], in which the contracted form is unsyllabifiable with the preceding /t/). This narrowed the envelope of variation to only those cases in which contraction was phonotactically licit. (Tokens of *is* and *has* after *it* are nevertheless included in this data set, because contraction in those cases is phonotactically licit.)
- Auxiliaries are found to **pattern in two groups**: those that contract at a rate of $\geq 90\%$ (*has, have, is, will*), and those that contract at a rate of $\leq 50\%$ (*had, would*).

- Possible explanations of this split in contraction rate:

- Word frequency: This could explain *is* and *had*, but not the other auxiliaries:

Approximate frequency per million words in Switchboard²: *is* (20,000) > *would* (3300) > *will* (2000) > *have* (1800) > *has* (900) > *had* (300).

Contraction rate: *is* (99%) > *have, will, has* (90%) > *would* (50%) > *had* (34%)

- Some phonological factor that makes /d/ less likely to attach to a host?

- A $[\pm\text{past}]$ condition on the variable process of contraction?

→ It's difficult to distinguish between these last two options. I'll be using $[\pm\text{past}]$ as a placeholder for a more definitive analysis.

3.2. Distribution of forms after noun phrase subjects

- Contraction is **much more restricted after noun phrases** than after pronouns:

(4) **Sue'll* [sul] (cf. *you'll* [jul])

(5) **three've* [θriv] (cf. *we've* [wiv])

- That is, **even where a contracted form would be phonotactically licit, it does not surface**. The exception to this is the /z/-final auxiliaries *is* and *has*, which contract freely (e.g. *Sue's, tree's...*). This fact has caused some previous researchers (e.g. Kaisse (1983)) to stipulate that the process of contraction does not apply to non-*is/has* auxiliaries after noun phrase subjects.
- The failure of non-*is/has* auxiliaries to contract after noun phrases is borne out by the data. However, this is not to say that only full forms of auxiliaries surface where contraction is disallowed. On the contrary, the data reveal that **intermediate forms (IFs) are prevalent** (Figure 2) and must be accounted for.

²Both full and contracted forms of each auxiliary are included in these counts. Because Switchboard is not part-of-speech-tagged, there are cases in which a full or contracted auxiliary is indistinguishable from another, unrelated form (e.g. 's 'is' = 's [genitive marker]; *have* [auxiliary] = *have* [main verb]). In such cases, frequencies were approximated by pulling out 100 tokens at random of the ambiguous item, counting the proportion of relevant to irrelevant hits in that set of 100, and then using that ratio to estimate the number of relevant hits in the entire corpus.

- These IFs, in principle, have **two plausible sources**:

- (a) A low-level process of *h*-deletion that operates on underlyingly full forms (presuming earlier destressing and vowel reduction; see Kaisse, 1985): /hæd/ → [hǎd] → [hǎd̥] → [ǎd̥]
- (b) A repair process that inserts [ə] between two consonants that cannot be syllabified (similar to that proposed in Sells, 1983): it /d/ → it [əd] ‘it had’

- Since IFs thus could come from either underlyingly full or underlyingly contracted forms, **they obscure the underlying full/contracted distribution**. But they can be **reclassified** as being due to either (a) or (b), hence giving us a picture of the underlying distribution (Figure 3).

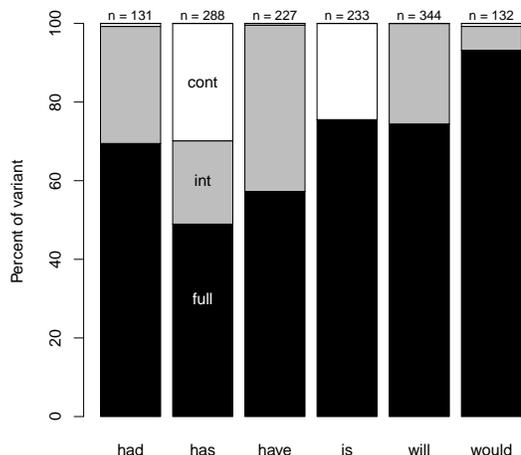


Figure 2: Surface distribution (observed) after noun phrase subjects.

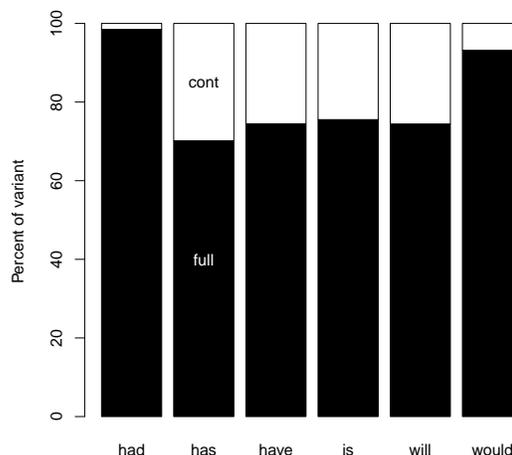


Figure 3: Underlying distribution (inferred) after noun phrase subjects.

- This reclassification hinges on the following assumptions:

1. IFs of *h*-initial auxiliaries (i.e., *had*, *has*, *have*) come from (a). There is no analogous process of *w*-deletion to create IFs from full forms of *will*, *would* (Kaisse, 1985).
2. IFs of auxiliaries that can't undergo close attachment to their noun phrase host (i.e., *had*, *have*, *will*, *would*) come from (b). Close attachment may fail for phonotactic reasons (i.e., because the resultant consonant cluster would be unsyllabifiable, as exemplified by *it'd* in (b) above) or because of these auxiliaries' general failure to contract after a noun phrase even where it would be phonotactically licit (as exemplified by **Sue'll*, **three've* in (4–5) above).

- IFs of *has* can thus be reclassified as underlyingly full (i.e., they come from (a)). (b) is not a plausible source of these IFs because *has* can contract after noun phrases with no need for a phonological repair.³
- IFs of *will* and *would* can be reclassified as underlyingly contracted. There is no process of *w*-deletion that could derive them from underlyingly full forms; thus, they must be the result of underlyingly contracted forms that fail to syllabify with their host.
- IFs of *have* and *had* are ambiguous between underlyingly full and underlyingly contracted, as both (a) and (b) are plausible sources of these forms. They were reclassified as follows:

- Assuming that the [±past] condition on contraction observed after pronouns holds after noun phrases as well, *have* must contract underlyingly at the same rate as other [-past], while *had* must contract at a lower rate. *h*-deletion is assumed to hold steady across all auxiliaries: this low-level process is not expected to show word-specific behavior.

³The one environment in which contracted *is* and *has* do require (b) in order to surface—namely, after sibilant-final noun phrase subjects—has been excluded from these graphs.

– Given these assumptions, the rates of contraction and *h*-deletion exhibited by the other auxiliaries can be applied to *have* and *had*, to calculate a predicted distribution of IFs, as follows:

- * *Will* displays a contraction rate of 26%⁴. *Has* displays a rate of *h*-deletion of 30%. Applying these two rates to the 227 tokens of *have* yields a predicted 109 intermediate and 118 full forms, a distribution that does not differ significantly from what is observed ($\chi^2 = 2.541$, $p = 0.111$).
- * Applying these rates to the 131 tokens of *had* yields a predicted 63 intermediate forms and 68 full forms. This, however, is significantly different from what is observed ($\chi^2 = 16.176$, $p < .001$). Contraction must apply at a lower rate to *had* than to the non-past auxiliaries.

- Given judicial treatment of ambiguous intermediate forms, then, **the [\pm past] condition resurfaces**.
- Additionally, **contraction can be said to apply to all auxiliaries after noun phrases**, despite its failure to surface as such in the majority of cases. It does, however, occur at a much lower rate than it does after pronouns.

4. Conclusions

- Recognizing **two stages of variable processes** uncovers a single underlying pattern. These stages are:
 1. Some higher-level alternation of full and contracted forms, which is
 - conditioned by auxiliary tense (past vs. non-past)
 - conditioned by subject type (pronoun vs. noun phrase)
 2. Subsequent low-level phonetic/phonological processes, namely
 - a \emptyset -insertion repair process, which applies categorically to contracted forms that cannot closely attach to their host
 - a process of *h*-deletion, which applies variably to *h*-initial full forms
- More generally, contraction provides evidence that **surface patterns of variation can provide a window into the nature of complex morphosyntactic phenomena**.
- And, conversely, proposing **variation at multiple stages of the derivation** helps explain the observed surface patterns.

5. Extensions

Future work will focus on:

- Contraction of *does* and *did* in *wh*-questions as a potential window into the [\pm past] condition
- Whether the process of contraction is best formalized as a suppletive allomorphic alternation or the output of a morphophonological process
- How best to incorporate rightward effects, e.g. the failure of contraction to occur before a gap (see, e.g., King, 1970, Inkelas and Zec, 1993)

References

- Godfrey, John J., Edward C. Holliman, and Jane McDaniel. 1992. SWITCHBOARD: Telephone speech corpus for research and development. In *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing, Volume 1*, 517–520.
- Inkelas, Sharon, and Draga Zec. 1993. Auxiliary reduction without empty categories: A prosodic account. In *Working Papers of the Cornell Phonetics Laboratory 8*, ed. Corinne Moore and Ann Bradlow, 205–253.
- Kaisse, Ellen M. 1983. The syntax of auxiliary reduction in English. *Language* 59:93–122.
- Kaisse, Ellen M. 1985. *Connected Speech: The Interaction of Syntax and Phonology*. New York: Academic Press.
- King, Harold V. 1970. On blocking the rules for contraction in English. *Linguistic Inquiry* 1:134–136.
- Labov, William. 1969. Contraction, deletion, and inherent variability of the English copula. *Language* 45:715–762.
- McElhinny, Bonnie S. 1993. Copula and auxiliary contraction in the speech of White Americans. *American Speech* 68:371–399.
- Sells, Peter. 1983. Juncture and the phonology of auxiliary reduction in English. In *University of Massachusetts Occasional Papers in Linguistics 8*, ed. Toni Borowsky and Daniel Finer, 76–105.
- Zwicky, Arnold M. 1970. Auxiliary reduction in English. *Linguistic Inquiry* 1:323–336.

⁴This is not significantly different from the 30% contraction rate displayed by *has* ($\chi^2 = 1.280$, $p = 0.258$).