

Contraction Beyond the Copula

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 well-studied in sociolinguistics, with scholarship consisting mainly of a long tradition of studies of the contraction (and deletion, in African-American English) of the **copula** (*is, are*), initiated by Labov (1969).
- However, the **contraction of other auxiliaries** has received surprisingly little attention in quantitative studies. The one exception to this is McElhinny (1993), who did examine the contraction of auxiliaries other than the copula. The current work, a corpus-based study of contraction, seeks to supplement McElhinny's study by using a larger body of data and by permitting a wider envelope of variation (see Section 4.1, 2nd bullet). The questions at issue in this study are as follows:

Q1: What surface patterns do the data display? At what rate(s) does contraction occur in natural speech?

Q2: What factors condition this variation? This study focuses specifically on whether there is any effect of auxiliary identity or subject type.

Q3: What bearing do these findings have on how contraction must be represented in the grammar? Specifically, can it be proposed that all auxiliaries are subject to a single rule of contraction (as Rickford et al. (1991) proposed for *is* and *are*)? Or, are multiple contraction processes needed to explain the data?

- The following findings will be presented and discussed in Section 4:
 - After pronoun subjects (henceforth “post-pronoun”), auxiliaries display what appears to be a past vs. non-past split in contraction rate.
 - After full noun phrase subjects (“post-NP”), this past vs. non-past split is not evident in the surface data.
 - However, independent low-level processes subsequent to the process of contraction may simply be obscuring an underlying past vs. non-past split in the post-NP environment.
- More generally, these findings provide a reminder that **surface patterns may be explained by allowing variation at multiple stages of a derivation.**

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. The Process of Contraction

- Labov (1969) analyzes contracted auxiliaries as the output of a cascade of phonological rules: Vowel Reduction (to [ə]), /h/- and /w/-Deletion¹, and [ə]-Deletion. The application of these rules outputs a single consonant. Subsequent sociolinguistic work has adopted this analysis.
- Outside of sociolinguistics, however, contraction is typically analyzed as conditioned allomorphy: contracted forms are clitic allomorphs of full forms (e.g., Kaisse (1983), Inkelas and Zec (1993), Anderson (2008)).
- Kaisse (1985) provides a number of convincing arguments in favor of the morphological analysis over the phonological one. I follow her here in adopting an analysis of contraction by which full forms alternate with contracted ones in a process that occurs prior to the phonology.

¹Labov notes that contraction of *will, would* may be instead a “lexical alternation,” as /w/-deletion does not occur elsewhere in English.

4. Results and Discussion

4.1. Classification of the Dependent Variable

- 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)
- The existence of “intermediate” forms has been only briefly acknowledged in the literature. McElhinny (1993) explicitly omits intermediate forms from her study, defining contraction as the attachment of a single consonant to its host with no vowel between the two.² However, preliminary coding of auxiliary realization revealed that intermediate forms are prevalent; thus, I retain them as valid variants.

4.2. Factors Conditioning Auxiliary Realization

Q1: What surface patterns do the data display?

- Auxiliaries were examined after pronoun and full noun phrase subjects. The pronoun subjects considered were only those ending in a vowel (so, *I, he, they*, etc, but not *it*) unless the auxiliary was /z/-final (*has, is*), in which case *it* was included along with the vowel-final pronouns. In other words, auxiliaries after *it* were omitted where they could not fully contract without a schwa (i.e. *it'll, it'd* as opposed to *it's*). Henceforth, “pronoun subjects” should be taken to mean this set of pronouns investigated.

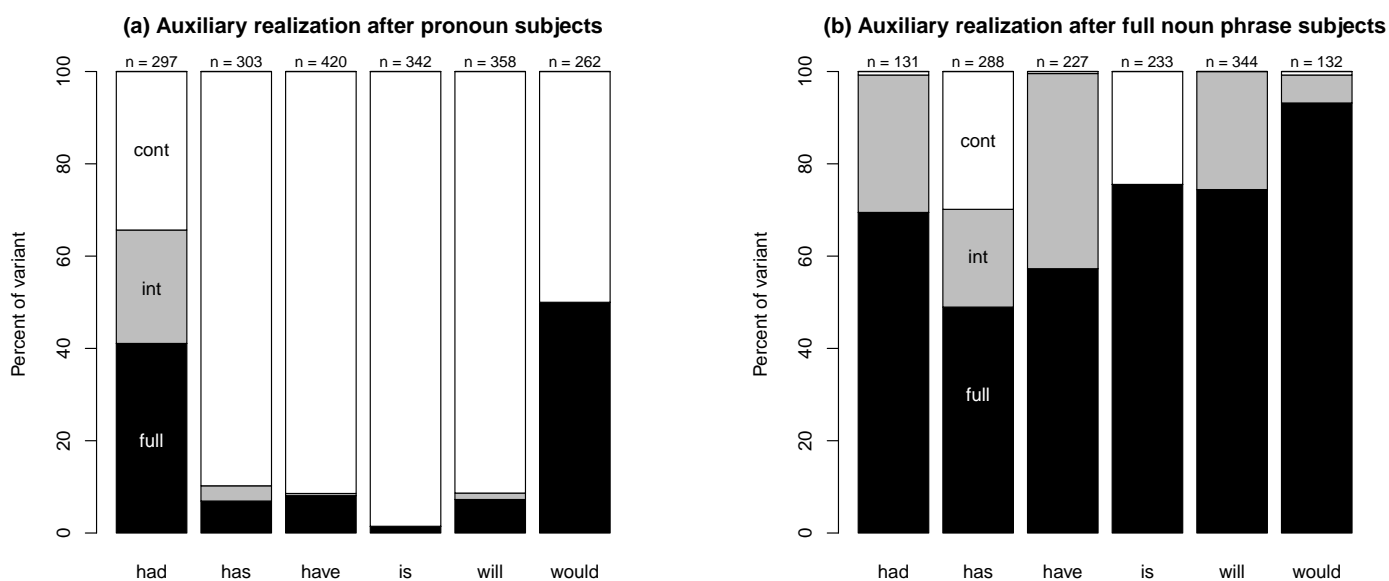


Figure 1: Distribution of forms after pronouns (a) and noun phrases (b). Black = full, gray = intermediate, white = contracted.

- Figure 1 shows a much higher rate of contraction after pronoun subjects than after noun phrase subjects, a result familiar from work on the copula. But the post-pronoun data reveal that **not all auxiliaries contract at the same rate**. Rather, auxiliaries **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*).

²McElhinny (p. 392) suggests, though, that future work reconsider this definition of contraction.

Q2: What can this split in contraction rate be attributed to?

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

Approximate frequency per million words in Switchboard³: *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 past] condition on the variable process of contraction?
→ It's difficult to distinguish between these last two options. I'll be using [\pm past] as a placeholder for a more definitive analysis.

Q3: Is this split in evidence after noun phrases, as well?

- ... or are post-pronoun and post-NP contraction processes differently conditioned, providing an argument for there being separate, environment-specific processes?
- The sharp past vs. non-past split in evidence in Figure 1a is clearly not apparent in Figure 1b. But before writing off a [\pm past] condition after noun phrases, the numerous intermediate forms in Figure 1b must be accounted for. There are **two plausible sources of intermediate forms**:

(a) A low-level process of *h*-deletion that operates on underlyingly full forms (presuming earlier destressing and vowel reduction; see Kaisse, 1985): /hæd/ → [hæd] → [həd] → [əd]

This process affects other /h/-initial function words elsewhere in English: e.g. *he, himself, her...*

(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'

This repair process occurs wherever a contracted form cannot surface as-is. Namely:

- Where the contracted form and the preceding consonant would make a phonotactically illicit cluster: e.g. *it'd, it'll*
- Where the auxiliary is non-/z/-final and the host is a noun phrase: e.g. **Sue'll [sul], *three've [θriv]*

- 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 2).

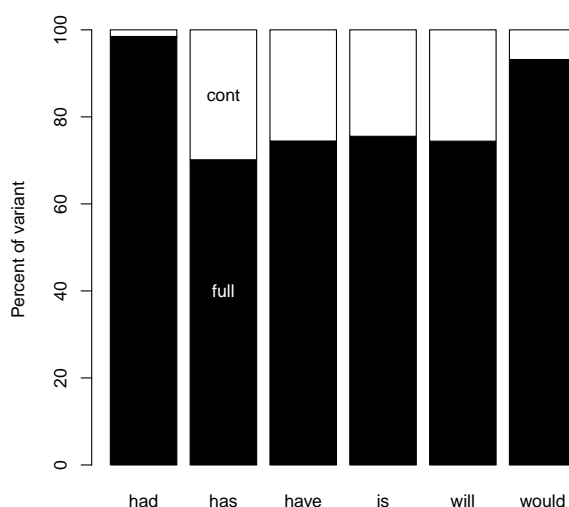


Figure 2: Underlying distribution (inferred) after noun phrase subjects.

³Both 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.

- 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 be syllabified with their noun phrase host (i.e., *had*, *have*, *will*, *would*) come from (b).
- 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.⁴
- 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 [\pm 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.
 - 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%.⁵ *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.**

5. Conclusions

- The surface realization of auxiliaries can be attributed to two stages:
 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 a reminder that **surface data cannot always be taken at face value**. In this case, surface data belie an underlying unity of process across subject types, which has been obscured by low-level phenomena.

⁴The one environment in which contracted *has* **would** require (b) in order to surface—namely, after sibilant-final noun phrase subjects—has been excluded from these graphs.

⁵This is not significantly different from the 30% contraction rate displayed by *has* ($\chi^2 = 1.280$, $p = 0.258$).

6. Extensions

Future work will investigate:

- Contraction of *does* and *did* in *wh*-questions as a potential window into the [\pm past] condition (e.g. *How's that sound?*, *What's it look like?*, *Why'd he do that?*)
- The exact nature of the contraction process, incorporating its failure to occur before gaps (as observed by Labov (1969) and many other contraction researchers outside sociolinguistics)
- The external constraints on contraction, probing preliminary findings from Switchboard, plotted in Figures 3 and 4

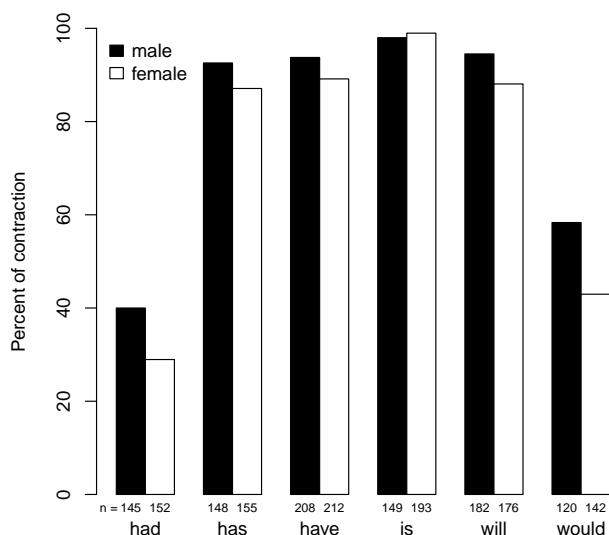


Figure 3: Percent of contracted forms per auxiliary, post-pronouns, by sex. *Had*, *will*, and *would* show significant ($p < .05$) sex differences.

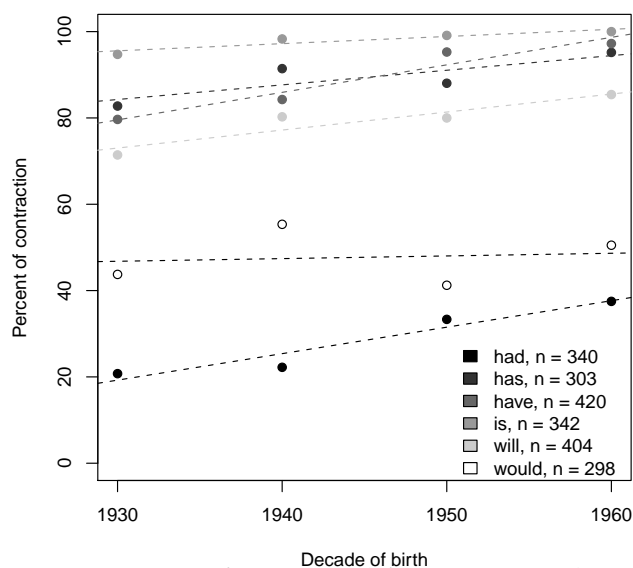


Figure 4: Percent of contracted forms per auxiliary, post-pronouns, by decade of birth. All auxiliaries show a significant age effect except *would*.

References

- Anderson, Stephen R. 2008. English reduced auxiliaries really are simple clitics. *Lingue e Linguaggio* 7:169–186.
- 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.
- 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.
- Rickford, John R., Arnetha Ball, Renée 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* 3:103–132.
- 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.