

English Auxiliary Contraction as a Two-Stage Process: Evidence from Corpus Data

Laurel MacKenzie
University of Pennsylvania

1. Introduction

English auxiliary contraction has a rich history in the linguistic literature, with analyses spanning four decades' worth of work (e.g. Zwicky, 1970, Bresnan, 1978, Kaisse, 1983, Inkelas & Zec, 1993, Anderson, 2008). However, with the exception of work in the sociolinguistic domain (e.g. Labov, 1969, McElhinny, 1993, Walker & Meechan, 1999), work on contraction has relied exclusively on authors' intuitions, with no examination of the patterns observed in natural speech. As a result, descriptions of this phenomenon have failed to take into account the factors that condition its variation, which may have implications for where in the grammar it is situated. The sociolinguistic work on contraction, by contrast, is founded on empirical data, but is incomplete for another reason: the low frequency of some auxiliaries in natural speech has led many researchers to focus exclusively on contraction of the copula, overlooking any questions about unity of process that are raised by the contraction of other auxiliaries.

For a number of reasons, the time is now right to address these concerns by revisiting this variable. For one, the advent of massive speech corpora has facilitated the study of low-frequency morphosyntactic phenomena. Additionally, a detailed study of contraction can contribute to recent discussions concerning the place of variation within the grammar (see Adger, 2006:506 for a list of relevant work), by adding to our knowledge of which linguistic elements can vary and which elements can condition that variation.

To that end, this paper describes the results of a corpus study of English auxiliary contraction and their implications for the place of variation in a grammatical architecture. Specifically, I argue that the quantitative patterns displayed by this variable lend themselves to a two-stage analysis of contraction, as an allomorphic alternation followed by the application of low-level phonetic or phonological processes. I support this analysis with data showing an effect of subject length on contracted forms, a finding which raises important questions about the grammatical locus of such a conditioning factor. This study thus underscores the importance of quantitative data in theoretical analysis.

The remainder of the paper is structured as follows. Section 2 describes the phenomenon of auxiliary contraction and surveys previous literature on this alternation. Section 3 describes the methodology of the corpus study. Section 4 presents the pattern of variation found in the corpus data and proposes the two-stage analysis to account for it. Section 5 provides support for this two-stage analysis with a study of the effect of subject length on contraction rate, and discusses the implications of the observed subject length effect for a grammatical architecture. Section 6 concludes.

2. Background

2.1. *The phenomenon*

The phenomenon under study in this paper is the variable alternation between full and contracted forms of the auxiliaries *has*, *have*, *is*, and *will*.¹ The precise phonological shape of an auxiliary's "contracted form" will be outlined below; the basic gist of the phenomenon is that a form of the auxiliary

*Thanks to David Embick, Anthony Kroch, William Labov, and audiences at NWAV 39, the LSA 2011 annual meeting, and the University of Pennsylvania.

¹Though the auxiliaries *had* and *would* contract as well, MacKenzie (to appear) finds that their rate of contraction after noun phrases, the environment primarily at issue in this paper, is too low for any meaningful effects to be apparent. Likewise, the auxiliaries *am*, *are*, *does*, and *did* contract too, but after a set of environments that is too limited to allow in-depth study.

with all its phonological material alternates with one that is phonologically reduced and cliticized to the preceding word.

2.2. Previous work and current analysis

While contraction has been the subject of a large body of work, much of it has focused on where contracted forms *cannot* occur (for instance, their failure to surface before gaps), rather than what has occurred when they *do* surface. Nonetheless, a few such analyses have been put forward.

The earliest work (e.g. Zwicky, 1970, Sells, 1983) treated contracted forms as the output of a cascade of phonological processes: deletion of an auxiliary’s initial /h/ or /w/, deletion of its vowel, and resyllabification of the remaining consonant with the preceding word. An alternative analysis of contraction as an alternation between full and contracted allomorphs was first proposed by Kaisse (1983). Arguing against an analysis under which contracted forms are derived from full by the phonology, Kaisse cites the fact that the rules of /h/- and vowel-deletion that Zwicky makes use of are active only at high rates of speech, while contraction can occur at all speaking rates; additionally, she points out that a phonological account of *will* and *would* contraction requires a rule of /w/-deletion that is not attested elsewhere in the language (cf. the failure of *was* and *were* to contract). Subsequent work (e.g. Inkelas & Zec, 1993, Close, 2004, Anderson, 2008) adopts this allomorphic analysis, proposing that a more phonologically full form of each auxiliary alternates with a more phonologically reduced one.

Any analysis of contraction must also account for the mechanisms that condition the insertion of the contracted allomorph and join it together with its host in a single syllable. To accomplish this, Kaisse proposes a variable morphosyntactic process that adjoins an auxiliary to its host, feeding insertion of the contracted form of the auxiliary in question. When this adjunction does not occur, an auxiliary’s full form is inserted by default. I further assume that in those cases where adjunction has occurred, a subsequent process applies to enact some sort of “close phonology” between the host and auxiliary (on which more in Section 4.1). This brings the host and the contracted auxiliary into a single syllable and conditions the application of any vowel changes, such as those found when *will* contracts to pronouns. A sample derivation, in which variable adjunction has applied, is given in (1). The “contracted form” here is a single consonant with no vocalic material; this will be discussed further in Section 3.2.

- (1) Contraction: sample derivation²
- | | | | |
|----------------------------------|---|-----------------------------|---------------------|
| 3.sg \wedge T _[fut] | → | [[3.sg]T _[fut]] | adjunction |
| | | [[fɪ]l] | allomorph selection |
| | | [fɪl] _σ | close phonology |

Henceforth, I will use the term “adjunction” to refer to the rule of host–auxiliary adjunction that, under this analysis, feeds the insertion of contracted forms. I reserve the term “contraction” as a theory-neutral description of the alternation between full forms and forms that are missing phonological material.

3. Methodology of the corpus study

3.1. The corpus

Data for this study come from Switchboard (Godfrey et al., 1992), a corpus of short telephone conversations between strangers on assigned topics. Switchboard comprises 240 hours of speech (3 million words) by 542 unique speakers. A total of 3611 tokens of the four auxiliaries were coded.

3.2. Auxiliary forms

A preliminary study of the data found that, rather than dividing neatly into “full” and “contracted” forms, auxiliaries in fact surface in *three* distinct phonological shapes, as outlined in (2)–(4). Auxiliaries were thus coded according to this three-point scale. The third form was called “intermediate” for coding purposes, given that it is phonologically in between full and contracted forms. The term “intermediate” is used here only as a description of this particular form; a proper analysis of intermediate forms will be provided in Section 4. It is these forms that will necessitate the two-stage model of contraction.

²The notation used here is that of Embick (2007).

- (2) **Full:** audible initial consonant, vowel of any quality³
- | | |
|-------------------------------|-------------------------------|
| a. <i>has</i> : [hæz], [həz] | c. <i>is</i> : [ɪz], [əz] |
| b. <i>have</i> : [hæv], [həv] | d. <i>will</i> : [wɪl], [wəl] |
- (3) **Intermediate:** no initial consonant but audible vowel⁴
- | | |
|-----------------------|-----------------------|
| a. <i>has</i> : [əz] | c. <i>will</i> : [əl] |
| b. <i>have</i> : [əv] | |
- (4) **Contracted:** no initial consonant, no vowel; a single consonant that forms a syllable with its host and, in the case of *is* and *has*, assimilates to it in voicing
- | | |
|---|--|
| a. <i>has</i> : [z] (e.g. it's [ɪts], Jimmy's [dʒɪmɪz]) | c. <i>is</i> : [z] (e.g. it's [ɪts], Jimmy's [dʒɪmɪz]) |
| b. <i>have</i> : [v] (e.g. I've [aɪv], we've [wɪv]) | d. <i>will</i> : [l] (e.g. you'll [jʊl], we'll [wɪl]) |

3.3. Defining the envelope of variation

As noted in Section 2.2, contracted forms are restricted to a subset of environments in which auxiliaries can surface. Those environments in which choice of auxiliary shape is categorical were **omitted** from study, as follows: those in which an auxiliary precedes a movement (5) or a deletion (6) site, comparative sub-deletion constructions (7), pseudo-cleft constructions (8), and *thing-is* constructions (9). Though *is* is the only auxiliary that surfaces in many of these environments, in those cases where another auxiliary may also occur (the prime examples below), it similarly shows a categorical choice of form: neither intermediate nor contracted is grammatical.

- (5) I wonder where Gerard {[ɪz] / *[z]} __ today. (King, 1970)
- (6) I'm as tall as Bill {[ɪz] / *[z]} __. (Sells, 1983)
- (6') I've been there, but no one else in my family {[hæz] / *[əz] / *[z]} __.
- (6'') I haven't been there, but some of my friends {[hæv] / *[əv]} __.
- (6''') I'll go, but I doubt John {[wɪl] / *[əl]} __.
- (7) Marie's a better scientist than Pierre {[ɪz] / *[z]} an __ engineer. (Anderson, 2008)
- (8) What I wonder {[ɪz] / *[z]} whether we'll win. (Kaisse, 1983)
- (9) The thing {[ɪz] / *[z]}, we got to have lunch together. (Massam, 1999)

However, there is another set of environments in which contracted forms are likewise illicit, but variation in auxiliary shape still occurs, as intermediate forms are nonetheless grammatical. These environments are cases of *have* and *will* following noun phrase subjects (10–11).

- (10) Many have attempted the trek, but only three {[hæv] / [əv] / *[v]} made it to the top.
- (11) Sue {[wɪl] / [əl] / *[l]} be there by five.

Because some degree of variation in auxiliary realization can occur in these environments, they were **retained** in the current study. The question of why contracted forms (i.e., those consisting of only a single consonant, as defined above) are illicit in these environments will be revisited in Section 4.1.

4. Findings and analysis

The first finding of note is that, as reflected in the coding scheme and evidenced in Figure 1, intermediate forms of *has*, *have*, and *will* surface with some frequency after noun phrase subjects. Given

³I assume that the reduction of an auxiliary's underlying vowel to [ə] is the result of a separate process of fast-speech Vowel Reduction that does not interact with the processes governing contraction (contra earlier phonological analyses of the phenomenon). See Kaisse (1985) for more on fast-speech Vowel Reduction.

⁴No intermediate form of *is* distinct from its full form was coded, as *is* has no initial consonant to lose.

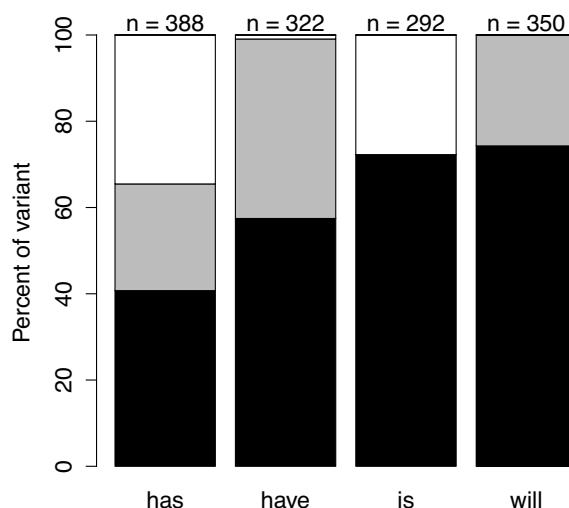


Figure 1: Distribution of forms after noun phrase subjects. Black = full, gray = intermediate, white = contracted.

that these forms are so well represented in natural speech, a thorough description of auxiliary contraction needs to account for them.

However, an analysis of intermediate forms is difficult to find in previous work. Kaisse (1983:98) notes the existence of “version[s] of the AUX that contain an initial schwa”; having noted in addition that “it is sometimes impossible to distinguish this allomorph from [...] regular, phonologically produced reduction,” she focuses throughout on *is* and *has* in order to sidestep this complication. Other work is contradictory on the subject of whether these intermediate forms are to be classified with an auxiliary’s contracted form: Anderson (2008) specifies that the alternative to the full form of an auxiliary is a single consonant, but then exemplifies this form with *Fred’ll*, where the auxiliary must clearly have a schwa; Inkelas & Zec (1993) denote the contracted form of each auxiliary as consisting of a single consonant, with the exception of *have*, whose contracted form they transcribe as [əv] without explanation. Given that these authors do not examine natural speech data, it may be unsurprising that they do not take the full range of auxiliary shapes into account. But even in sociolinguistic work on contraction, intermediate forms have been explicitly omitted (McElhinny, 1993).

The precise nature of intermediate forms is not a trivial question: it has implications for the nature of the mechanisms governing contraction. Specifically, it raises the possibility that the tripartite surface distinction between full, intermediate, and contracted reflects an underlying three-way competition for insertion, i.e., with one allomorph for each distinct surface form. If this is the case, the conditions governing the insertion of each form must be identified. Alternatively, it may be that surface intermediate forms can be explained away as the result of some phonetic or phonological processes, and that auxiliaries undergo an alternation that is simply bipartite, i.e., between two allomorphs.

In Section 4.1, I will argue in favor of the latter analysis. Specifically, I propose that intermediate forms are simply the artifact of low-level processes that occur after adjunction: a two-stage analysis of auxiliary realization. This analysis will be supported, in Section 5, by data from a study of the effect of subject length on contraction.

4.1. A two-stage model of contraction

The first clue toward the source of intermediate forms of the auxiliaries *has*, *have*, and *will* is in the distribution of each auxiliary’s variants. Specifically, *has* differs from *have* and *will* in permitting both contracted and intermediate forms after noun phrases: contracted *John[z]* and intermediate *John [əz]* are both acceptable. By contrast, contracted forms of *have* and *will* do not surface after noun phrases, even where they would be phonotactically licit: compare the ungrammatical *three’ve* *[θriv] and *Sue’ll* *[sul].

Based on this distinction, I propose that intermediate forms in the former and latter groups come

from different sources: that is, there are two distinct ways of deriving intermediate forms. Specifically, I will argue, intermediate forms in the first case are underlyingly full; intermediate forms in the second case are underlyingly contracted. I'll address these two categories in reverse order.

4.1.1. Intermediate forms from underlyingly contracted forms

A crucial point to recognize in our analysis of intermediate forms is that just because an auxiliary fails to surface in its contracted form in a particular environment does not mean that the process of morphosyntactic adjunction given in (1) has not occurred. Compelling evidence for this comes from the patterning of the auxiliary *will* after pronouns, depicted in Figure 2.

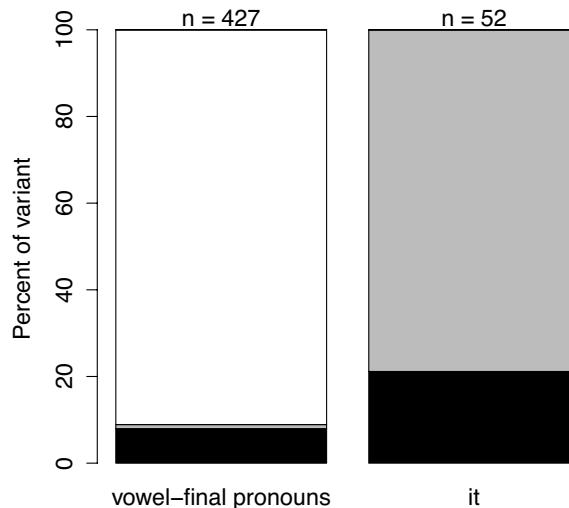


Figure 2: Distribution of forms after pronoun subjects, *will* only. Black = full, gray = intermediate, white = contracted.

After vowel-final personal pronouns (*he, she, I, etc.*), *will* surfaces in its contracted form 91% of the time. After the pronoun *it*, *will* surfaces 79% of the time in its intermediate form. Contracted forms of *will* are, of course, illicit after *it*, given English phonotactics. But the surface distribution strongly resembles complementary distribution of contracted and intermediate forms across the different phonological environments.

This complementary distribution can be explained by appealing to the model of contraction in (1), in which adjunction is followed by allomorph selection and close phonology. When adjunction of *will* to its host occurs, the single-consonant contracted allomorph will be inserted. Close phonology will subsequently try to apply, to bring the host and auxiliary together into a single syllable. It will succeed when the host is a vowel-final pronoun, but fail when the host is *it*. In this latter scenario, there is then a repair process, which I call Schwa Insertion, that inserts [ə] between the pronoun and the unsyllabified consonant. This explains why contracted [l] after vowel-final pronouns occurs at a rate comparable to that of intermediate [əl] after *it*: morphosyntactic adjunction and contracted-form insertion have applied regardless of host; the contracted form simply gets a different phonological shape on the surface after a consonant. We now have one of the two sources of surface intermediate forms: Schwa Insertion before contracted forms where close phonology cannot occur.

Taking this a step further, I propose that this analysis can be extended to other instances in which contracted forms fail to surface: that is, not only after *it*, but also after noun phrases. Forms like *three* [əv] and *Sue* [əl], which have no counterparts *[θriv] and *[sul], may again be the result of a contracted form failing to undergo close phonology, with a [ə] repair. An obvious question here is why close phonology would fail in these cases where it should be phonotactically acceptable. It may be related to the similar failure of contracted forms to surface after conjoined and embedded pronouns (Zwicky, 1970): *John and I'll* *[ajl] *be there* and *The guy sitting next to you'll* *[ju:l] *be there* are both illicit with close phonology of the contracted form, but acceptable with an intermediate form instead. If close

phonology is sensitive to the number of prosodic brackets separating host and auxiliary, for example, and noun phrases and embedded pronouns both have more prosodic brackets following them than do bare pronouns, it is reasonable that close phonology would be blocked in the former environments. Although a precise analysis of this awaits future work, what is crucial for our purposes is this: just because close phonology has failed does not mean that adjunction and contracted form insertion have not occurred.

The source of the intermediate forms discussed in this section thus stands as follows: they are contracted forms that have undergone Schwa Insertion where close phonology has failed, for simple phonotactic or other reasons.

4.1.2. Intermediate forms from underlyingly full forms

We have now accounted for post-noun phrase intermediate forms of *have* and *will*, but post-noun phrase intermediate forms of *has* still remain to be explained. If the same explanation were to hold for *has* as was provided for the other two auxiliaries, these intermediate forms would be generated from contracted forms that fail to undergo close phonology with their host, and subsequently undergo Schwa Insertion. However, contracted forms of *has* clearly **can** undergo close phonology with their host: hence the presence in the data of the single-consonant contracted form of *has*, at a rate of nearly 40%.⁵ Close phonology would have to be variable in order to account for this distribution.

Rather than attempt to argue for variable syllabification, I propose that intermediate forms of *has* have a different source altogether: they come instead from underlyingly full forms that have lost their initial consonant. A natural source of this initial consonant loss is Kaisse's (1985) fast-speech /h/-Deletion, which affects pronouns and function words in connected speech. Combined with another process of vowel reduction to schwa, also well-attested in function words, /h/-Deletion gives us another way of deriving intermediate forms, this time from full forms of /h/-initial auxiliaries. Intermediate forms of both *has* and *have*, then, can come from /h/-Deletion on full forms; this fact will come into play when we examine the effects of subject length in Section 5.

4.1.3. Summary

We now have two sources of surface intermediate forms: fast speech /h/-Deletion on full forms, and Schwa Insertion for syllabification purposes on contracted forms. This analysis of intermediate forms maintains an underlying bipartite distinction between full and contracted allomorphs, despite there being a tripartite distinction on the surface. It does so by making reference to two stages of processes: one, a morphosyntactic rule adjoining host and auxiliary; the other, a set of low-level phonetic and phonological processes that act on the allomorph whose insertion was conditioned by the first stage. Table 1 summarizes the source of each auxiliary's surface forms after noun phrase subjects.

	<i>is</i>	<i>has</i>	<i>will</i>	<i>have</i>
Underlying contracted forms surface as:	contracted	contracted	intermediate (from [ə]-Ins.)	intermediate (from [ə]-Ins.)
Underlying full forms surface as:	full	full & interm. (from /h/-Del.)	full	full & interm. (from /h/-Del.)

Table 1: Source of surface forms after noun phrase subjects.

Under the present analysis, all four auxiliaries under study are subject to the same rule of adjunction. Likewise, for all auxiliaries, this adjunction rule conditions insertion of a single-consonant contracted form, though how this contracted form is realized on the surface differs from auxiliary to auxiliary. Presumably, as a variable process, adjunction has some conditions on its application, meaning that its application will be disfavored in particular linguistic environments. In those environments where adjunction is disfavored, then, we should expect to see fewer surface realizations of underlying contracted forms. This was tested through the subject length study described in Section 5.

⁵Strictly speaking, there is one environment in which contracted forms of *has* **cannot** undergo close phonology with their host: namely, after sibilants. Sibilant-final noun phrase subjects have thus been omitted from study.

5. The effect of subject length

5.1. Background

The role of noun phrase length or weight has been investigated in studies of a number of variable phenomena, the majority of which involve movement of a long or heavy phrase to the end of a sentence (the phenomenon of “end weight”). Phenomena in which NP length or weight effects have been demonstrated include heavy-NP shift (Wasow, 1997), relative clause extraposition (Francis, 2010), the dative alternation (Wasow, 1997), the genitive alternation (Rosenbach, 2005), particle shift in verb-particle constructions (Wasow, 1997), and the ordering of elements in a binomial (Benor & Levy, 2006). Explanations of these end weight effects seem to fall into two camps: one camp describes them as stemming from a need to have as much time as possible to plan long and/or complex material (Wasow, 1997; Stallings & MacDonald, 2011); the other sees them as due to a tendency to minimize the domain between a verb and its constituents (e.g. Francis, 2010).

Where contraction is concerned, these proposals are not obviously relevant: regardless of whether a full or a contracted allomorph is used, the auxiliary is always immediately adjacent to its noun phrase host. Still, some indication that contraction may be sensitive to host heaviness has already been demonstrated in past work, in the much lower rate of copula contraction after full noun phrases than after pronouns first shown by Labov (1969). Given that pronouns are prosodically lighter than full noun phrases, it is conceivable that subject weight may be a conditioning factor on the morphosyntactic adjunction rule that governs the insertion of contracted forms. To that end, each token of a post-noun phrase auxiliary was coded for the number of orthographic words in its subject.

Much work has addressed the question of the appropriate measure of NP weight; number of syllables, prosodic words, orthographic words, discourse-new referents, syntactic nodes, and major phrase boundaries have all been investigated, and are frequently difficult to tell apart as they are highly correlated. Szmrecsányi (2004) concludes that the strong correlation between the various measures means that there is no reason not to use the one that is simplest to measure, that is, word count; Shih & Grafmiller (2011) come to a similar conclusion, finding syntactic node count to be the best predictor of dative and genitive alternations through several statistical tests, but word count to be a close second and a reasonable proxy, hence its use in the current study.

5.2. Findings

Figure 3 opposes, for each auxiliary, the hypothesized surface manifestation of its full form to the hypothesized surface manifestation of its contracted form. Each data point represents a single token, coded for the number of words in its subject. We begin with *is*, for which underlying forms are hypothesized to surface as-is. The plot demonstrates a clear effect of subject length on adjunction of this auxiliary: as subjects increase in length, contracted forms taper off.

In the case of *has*, contracted forms are opposed to full and intermediate forms, as intermediate forms are by hypothesis underlyingly full forms that have undergone /h/-Deletion. A pattern identical to that observed for *is* surfaces: contracted forms taper off with subjects of increasing length.

This same effect is also in evidence for *will*, where intermediate forms have been opposed to full forms, under the analysis that underlyingly contracted forms of *will* cannot surface as-is after noun phrases, so intermediate forms surface in their place.

However, the plot for *have*, in which intermediate forms have again been opposed to full, is a clear outlier in this set of four, with no tapering off of intermediate forms as subject length increases. This is in fact expected under the current analysis, which attributes intermediate forms of *have* to **two** sources: as shown in Table 1, contracted forms of *have* surface as intermediate after noun phrases since they fail to undergo close phonology with their host, but full forms of *have* may also surface as intermediate, as they are subject to /h/-Deletion. Intermediate forms of *have* are thus of ambiguous origin, and there is no way to separate their two sources on the surface.⁶ As a result, they fail to show the same subject length effect, with a number of them — by hypothesis, those that were full underlyingly — continuing to surface after long subjects. This plot, then, can be taken as clear evidence of a fundamental difference

⁶See MacKenzie (to appear) for an attempt to identify the sources of these ambiguous intermediate forms of *have* by applying the rates at which adjunction and /h/-Deletion apply to other auxiliaries.

between intermediate forms of *have* and those of *will*, and also as an indication that /h/-Deletion shows no effect of subject length, an unsurprising conclusion given its nature as a low-level phonetic process.

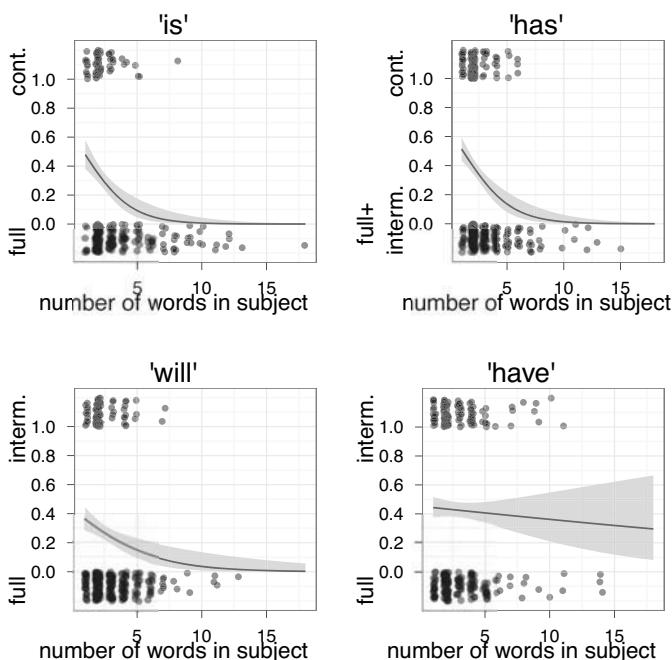


Figure 3: Distribution of surface forms by number of words in subject, showing decline in use of underlying contracted forms as subject length increases. “Cont.” = contracted, “interm.” = intermediate.

5.3. Implications

The plots in Figure 3 provide a clear confirmation of the analysis of contraction proposed here. Those surface forms that can be uniquely attributed to underlyingly contracted forms taper off as their subjects increase in length. But the finding that contraction should show an effect of subject length is in itself interesting, and becomes even more interesting when we observe that contracted forms of *is*, *has*, and *will* appear to exhibit a cut-off effect: in the plots above, no auxiliary shows contracted forms after subjects of more than eight words.⁷

This finding raises a number of questions about the precise nature of the subject length effect. For instance, why should the cut-off be at eight — is this effect somehow related to short-term memory capacity? What linguistic unit is actually driving the effect — might prosodic words be the operative unit, rather than orthographic words, or might duration or intonation contours be playing a role?

Also at issue is the question of where this eight-word effect should be localized in the model of contraction provided here. It would be simple to represent all factors that favor or disfavor the application of host–auxiliary adjunction on the adjunction rule itself. But incorporating the eight-word effect on our morphosyntactic rule of adjunction would mean that the grammar can be sensitive to numbers as precise as eight: an undesired consequence, as *categorical* grammatical operations that can count that high are unattested. Instead, it may be more appropriate to localize the eight-word cut-off in a system of “usage” distinct from the grammar, to allow for the fact that it is only *variable* alternations that may be susceptible to such conditioning. Future work will address this question in more detail, examining additional factors that are found to condition variable alternations but not categorical ones and determining whether they, too, are best seen as the purview of a different system.

⁷There is one exception to this finding, which has not been plotted, but it is a special case. The token, for which the subject is in bold, is as follows: *With desktop publishing and stuff, I think **a lot more magazines that are— that aim to a smaller market** [ə] be coming out.* The complete subject exceeds eight words in length, but contains a restart. The material after the restart does not exceed eight words.

6. Conclusion

This paper has reported on a quantitative corpus study of auxiliary contraction in English, the first such study to devote attention to what have been called “intermediate” forms. I have analyzed these intermediate forms as the output of two stages of processes: variable morphosyntactic adjunction which conditions insertion of either a full or a contracted form, then low-level phonetic and phonological processes which alter the phonological shape of that form. I have shown that, given judicious treatment of intermediate forms, a subject length effect appears in the data by which contracted forms are disfavored after longer subjects, with an apparent eight-word cut-off after which contracted forms categorically do not surface. This finding has subsequently raised a number of questions concerning where such an effect must be localized in a grammatical architecture, and whether some conditions on alternations should be treated as the purview of a system distinct from the grammar.

References

- Adger, David (2006). Combinatorial variability. *Journal of Linguistics* 42:3, 503–530.
- Anderson, Stephen R. (2008). English reduced auxiliaries really are simple clitics. *Lingue e Linguaggio* 7:2, 169–186.
- Benor, Sarah Bunin & Roger Levy (2006). The chicken or the egg? A probabilistic analysis of English binomials. *Language* 82:2, 233–278.
- Bresnan, Joan (1978). Contraction and the transformational cycle in English. Indiana University Linguistics Club, Bloomington, Indiana.
- Close, Joanne (2004). *English Auxiliaries: A Syntactic Study of Contraction and Variation*. Ph.D. thesis, University of York.
- Embick, David (2007). Linearization and local dislocation: Derivational mechanics and interactions. *Linguistic Analysis* 33:3-4, 303–336.
- Francis, Elaine J. (2010). Grammatical weight and relative clause extraposition in English. *Cognitive Linguistics* 21:1, 35–74.
- Godfrey, John J., Edward C. Holliman & Jane McDaniel (1992). SWITCHBOARD: Telephone speech corpus for research and development. *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing, Volume 1*, 517–520.
- Inkelas, Sharon & Draga Zec (1993). Auxiliary reduction without empty categories: A prosodic account. Moore, Corinne & Ann Bradlow (eds.), *Working Papers of the Cornell Phonetics Laboratory* 8, 205–253.
- Kaisse, Ellen M. (1983). The syntax of auxiliary reduction in English. *Language* 59:1, 93–122.
- Kaisse, Ellen M. (1985). *Connected Speech: The Interaction of Syntax and Phonology*. Academic Press, New York.
- King, Harold V. (1970). On blocking the rules for contraction in English. *Linguistic Inquiry* 1:1, 134–136.
- Labov, William (1969). Contraction, deletion, and inherent variability of the English copula. *Language* 45:4, 715–762.
- MacKenzie, Laurel (to appear). Locating linguistic variation: A case study of English auxiliary contraction. LaCara, Nick, Lena Fainleib & Yangsook Park (eds.), *Proceedings of NELS 41*.
- Massam, Diane (1999). *Thing is* constructions: The thing is, is what’s the right analysis? *English Language and Linguistics* 3:2, 335–352.
- McElhinny, Bonnie S. (1993). Copula and auxiliary contraction in the speech of White Americans. *American Speech* 68:4, 371–399.
- Rosenbach, Anette (2005). Animacy versus weight as determinants of grammatical variation in English. *Language* 81:3, 613–644.
- Sells, Peter (1983). Juncture and the phonology of auxiliary reduction in English. Borowsky, Toni & Daniel Finer (eds.), *University of Massachusetts Occasional Papers in Linguistics* 8, 76–105.
- Shih, Stephanie & Jason Grafmiller (2011). Weighing in on end weight. Paper presented at the 2011 annual meeting of the Linguistic Society of America.
- Stallings, Lynne M. & Maryellen C. MacDonald (2011). It’s not just the “heavy NP”: Relative phrase length modulates the production of Heavy-NP shift. *Journal of Psycholinguistic Research*.
- Szmrecsányi, Benedikt (2004). On operationalizing syntactic complexity. *Actes des 7es Journées internationales d’Analyse statistique des Données Textuelles*, 1031–1038.
- Walker, James A. & Marjory Meechan (1999). The decreolization of Canadian English: Copula contraction and prosody. Jensen, John & Gerard Van Herk (eds.), *Actes du Congrès annuel de l’Association canadienne de linguistique 1998/Proceedings of the 1998 Annual Conference of the Canadian Linguistic Association*, 431–441.
- Wasow, Thomas A. (1997). Remarks on grammatical weight. *Language Variation and Change* 9:1, 81–105.
- Zwicky, Arnold M. (1970). Auxiliary reduction in English. *Linguistic Inquiry* 1:3, 323–336.

Proceedings of the 29th West Coast Conference on Formal Linguistics

edited by Jaehoon Choi, E. Alan Hogue,
Jeffrey Punske, Deniz Tat,
Jessamyn Schertz, and Alex Trueman

Cascadilla Proceedings Project Somerville, MA 2012

Copyright information

Proceedings of the 29th West Coast Conference on Formal Linguistics
© 2012 Cascadilla Proceedings Project, Somerville, MA. All rights reserved

ISBN 978-1-57473-451-5 library binding

A copyright notice for each paper is located at the bottom of the first page of the paper.
Reprints for course packs can be authorized by Cascadilla Proceedings Project.

Ordering information

Orders for the library binding edition are handled by Cascadilla Press.
To place an order, go to www.lingref.com or contact:

Cascadilla Press, P.O. Box 440355, Somerville, MA 02144, USA
phone: 1-617-776-2370, fax: 1-617-776-2271, sales@cascadilla.com

Web access and citation information

This entire proceedings can also be viewed on the web at www.lingref.com. Each paper has a unique document # which can be added to citations to facilitate access. The document # should not replace the full citation.

This paper can be cited as:

MacKenzie, Laurel. 2012. English Auxiliary Contraction as a Two-Stage Process: Evidence from Corpus Data. In *Proceedings of the 29th West Coast Conference on Formal Linguistics*, ed. Jaehoon Choi et al., 152-160. Somerville, MA: Cascadilla Proceedings Project. www.lingref.com, document #2698.