

Conditions on auxiliary contraction as evidence for variation outside the grammar

Laurel MacKenzie
University of Pennsylvania

LSA, Portland, January 5, 2012

1

I'm reporting today on the findings from a large scale corpus study of auxiliary contraction in English that I've been carrying out. Particularly I'm going to be presenting data showing that

- Subject length effect on auxiliary contraction
- How is this represented grammatically?
- Revisiting questions of how variable phenomena are represented in grammar
- Proposing that some conditions on variation may not be grammatically encoded

2

the length of an auxiliary's subject has an effect on the rate of contraction of that auxiliary.

I'm then going to ask the question of how this length effect could be worked into speakers' grammatical representations of the contraction alternation, and, after revisiting some longstanding questions of how variable phenomena are represented in the grammar, I'm going to propose that some conditions on variation, such as the subject length effect I present, have their source outside of the grammar.

Auxiliary contraction

- While we were in town, **we'd** ride several miles away, like to school and the like, but when we were in the country, **we would** ride ten, twelve miles away. (sw_1181)
- Yeah, **Salzburg's** nice. **Austria's** nice. **Europe is** nice! (sw_1151)
- Oh, I'm sure **it's** been done. I'm sure **it has** been done. (sw_1060)

3

Just to make sure we're all on the same page, here are a few real-life examples of contraction, taken from the Switchboard corpus.

I'll be using this color-coding throughout, with pink to notate contracted forms and green to notate uncontracted ones -- and I'll talk a little more about what counts as contracted or uncontracted in a second.

In this talk I'll be restricting my data to

Auxiliary contraction

- Yeah, **Salzburg's** nice. **Austria's** nice. **Europe is** nice! (sw_1151)

is

has

will

4

tokens with full noun phrase --- non-pronoun --- subjects only.
And I'm focusing only on three auxiliaries: is, has, and will.

Research questions

- What conditions the variation between contraction and non-contraction?
- What can we infer from those conditions about the processes that generate contractions?

5

And to reiterate, the research questions at issue here are what conditions the variation between contraction and non-contraction – in other words, what governs whether or not a contracted form surfaces – and what we can infer from those conditions about the processes that generate contractions.

Data sources

- The Switchboard corpus (Godfrey et al., 1992)
- The Fisher corpus (Cieri et al., 2004)
 - 5-minute telephone conversations between strangers on a given topic
- The Philadelphia Neighborhood Corpus (Labov & Rosenfelder, 2011)
 - Sociolinguistic interviews carried out by trained fieldworkers

6

Data in the present study come from three sources.
And I've done all my coding based on audio files.

The dependent variable

Underlyingly: a bipartite allomorphic alternation

is /ɪz/ ~ /z/

has /hæz/ ~ /z/

will /wɪl/ ~ /l/

(MacKenzie 2011)

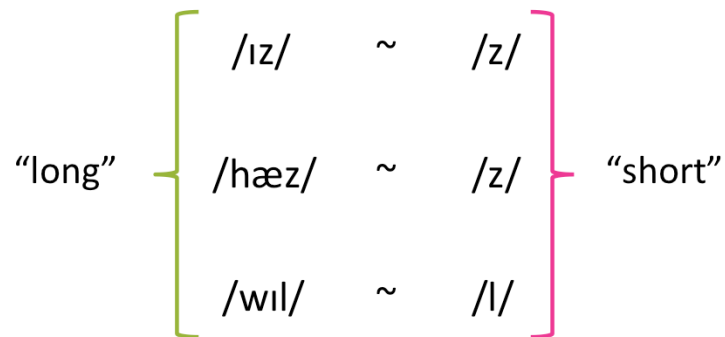
7

Before we get to the subject length effect, we need to take a closer look at the dependent variable.

In my previous work on contraction, I've analyzed the phenomenon as an underlying allomorphic alternation between two forms for each auxiliary.

The dependent variable

Underlyingly: a bipartite allomorphic alternation



(MacKenzie 2011)

8

For ease of exposition, I'm calling these forms on the left -- those with all their segmental material intact -- 'long forms', and those forms on the right -- those which consist of only a single consonant -- 'short forms'.

And when I use the word "contraction," I'm using it to mean "insertion of that short allomorph for some auxiliary."

The dependent variable

On the surface: numerous phonological shapes

<i>is</i>	[ɪz]	[əz]		[s], [z]
<i>has</i>	[hæz]	[həz]	[əz]	[s], [z]
<i>will</i>	[wɪl]	[wəl]	[əl]	[l]

(MacKenzie 2011)

9

Things aren't quite this simple on the surface: when you start listening to natural speech, you find that speakers produce a range of forms for each auxiliary -- more than the two that our underlying bipartite alternation would lead us to expect. Specifically, each auxiliary under study surfaces in a form that is phonologically faithful to the long allomorph -- so if an auxiliary has an initial consonant, this form retains it -- as well as a form that differs from that phonologically faithful one only in that it has a reduced vowel.

Those auxiliaries with initial consonants can also surface in a form that lacks them, and finally, each auxiliary can surface in a single-consonant form.

I've argued elsewhere that each of these forms can be attributed back to either the long or the short allomorph for that auxiliary. There's reason to believe that phonological and phonetic processes, such as vowel reduction and deletion of initial /h/, alter those underlying forms so that they surface in this array of different shapes, meaning that we can

The dependent variable

Surface forms can be reclassified:

	<u>Underlying long</u>		<u>Underlying short</u>	
<i>is</i>	[ɪz]	[əz]		[s], [z]
<i>has</i>	[hæz]	[həz]	[əz]	[s], [z]
<i>will</i>	[wɪl]	[wəl]	[əl]	[l]

(MacKenzie 2011)

10

reclassify these surface forms as stemming from either the underlying long allomorph, or the underlying short one, as follows.

So this reclassification of surface forms attributes each one to either the underlying long or the underlying short form, and this is how I've coded the data that I'm presenting today.

Where we have these forms on the left, we say contraction has not occurred; where we have these forms on the right, we say contraction has.

Now that we've sorted out the dependent variable, we can examine the

Subject length

Number of orthographic words in subject

4: I think the quality of TV's gone down (sw_1288)

6: The eastern half of the block has remained rather stable (PH79-3-6)

7: The one good thing about this country's you have a right to your opinion (fe_4912)

18: The idea, the merit, the thought of somebody who publicly was a member of the Ku Klux Klan is pretty repulsive (sw_1073)

11

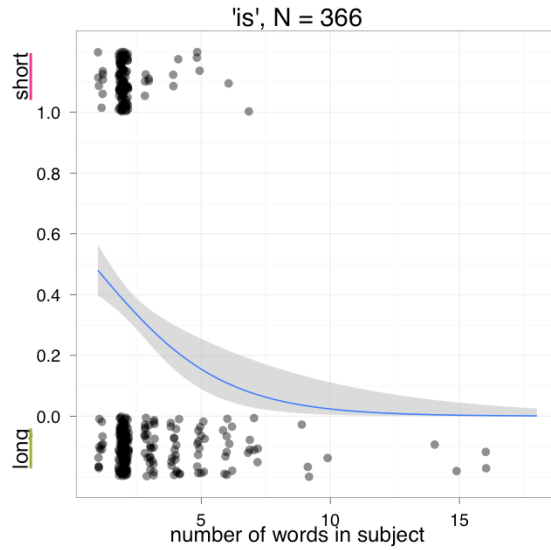
effect of subject length on contraction.

To start with, I measured subject length in the simplest way: by counting the number of orthographic words in an auxiliary's subject.

You can see here some examples of what sort of sentences you get for different possible subject lengths.

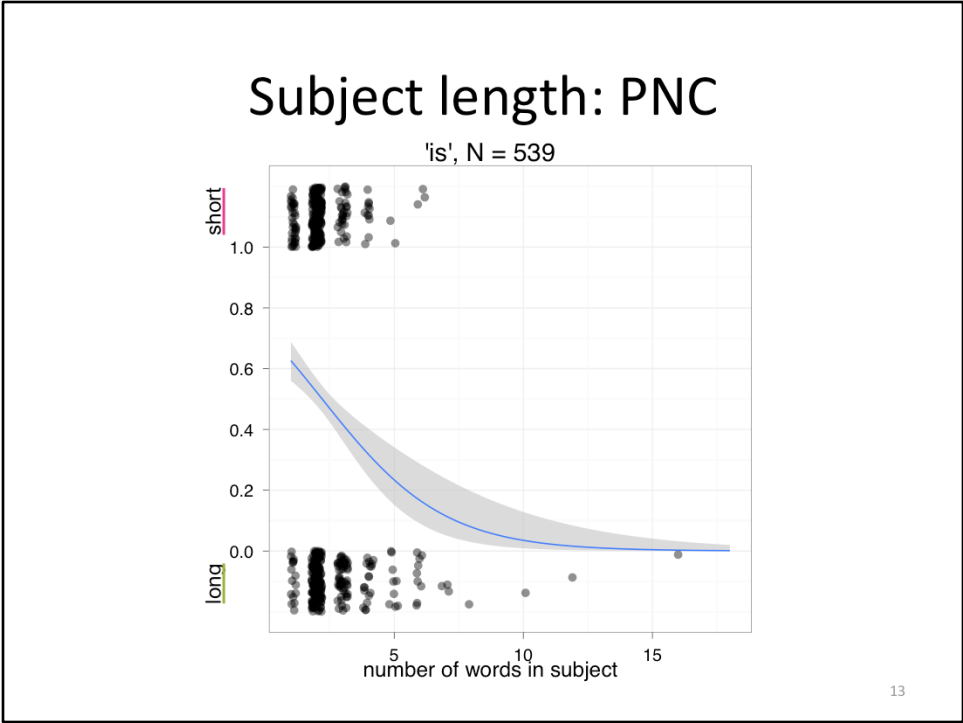
And the results were as follows:

Subject length: Fisher



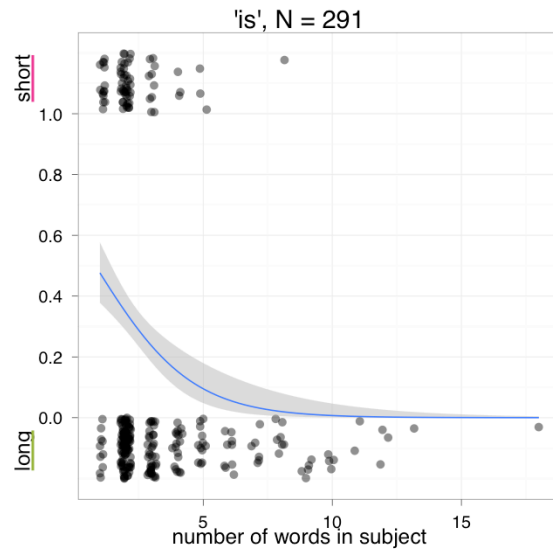
12

This plot shows data on the auxiliary 'is' from the Fisher corpus. Each dot represents a token, coded for whether its auxiliary surfaced as a form of the short allomorph or a form of the long one. The data are plotted by number of words in the subject on the x-axis, and you can see there's a downward trend by which short allomorphs taper off as subject length increases.



Here's the same thing from the Philadelphia Neighborhood Corpus

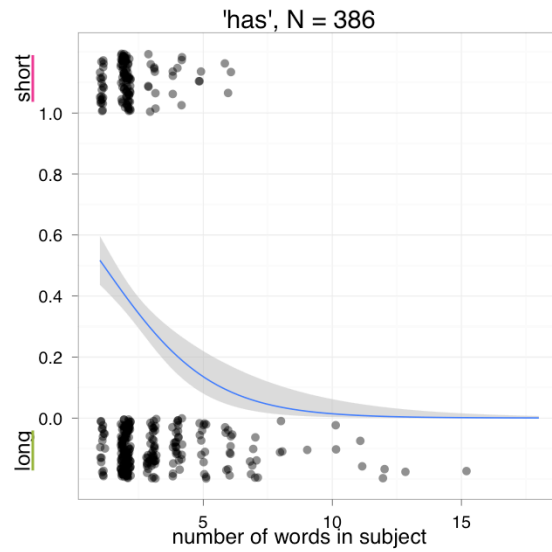
Subject length: Switchboard



14

and the same thing from Switchboard.

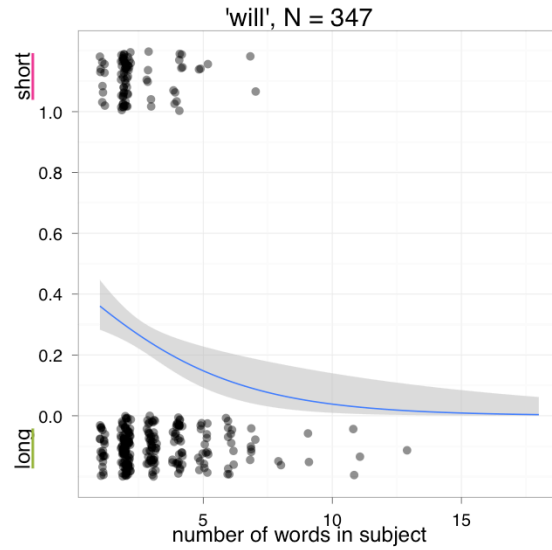
Subject length: Switchboard



15

This plot shows the auxiliary 'has' from Switchboard, again with all surface forms that can be attributed to the short allomorph opposed to all surface forms that can be attributed to the long one.

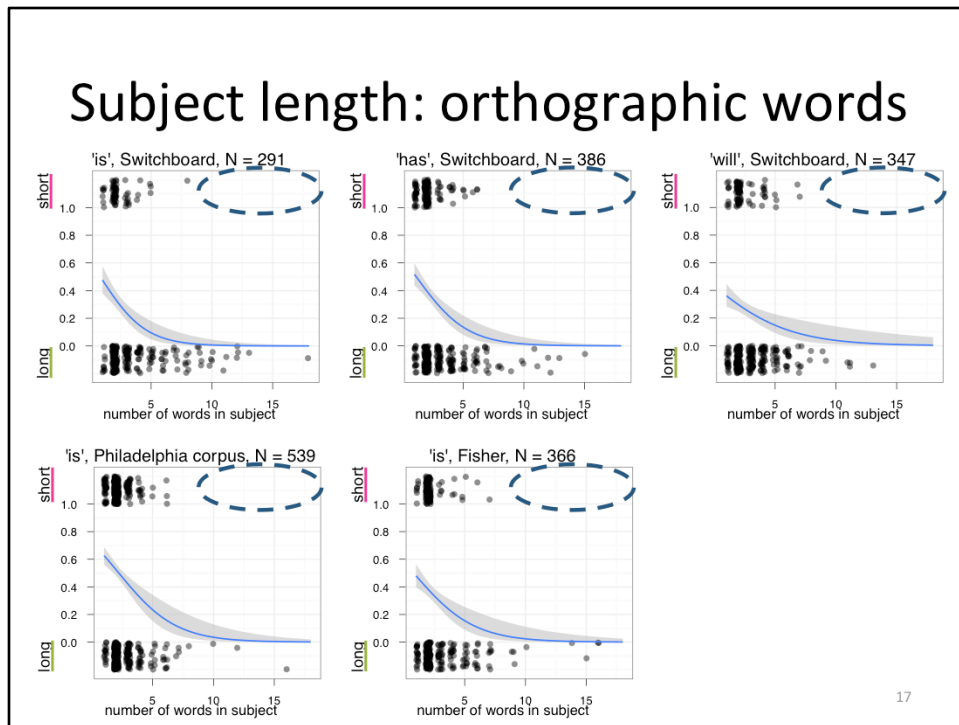
Subject length: Switchboard



16

And here's 'will' from Switchboard, again with all forms that can be attributed to the short allomorph opposed to all those that can be attributed to the long one.

Subject length: orthographic words



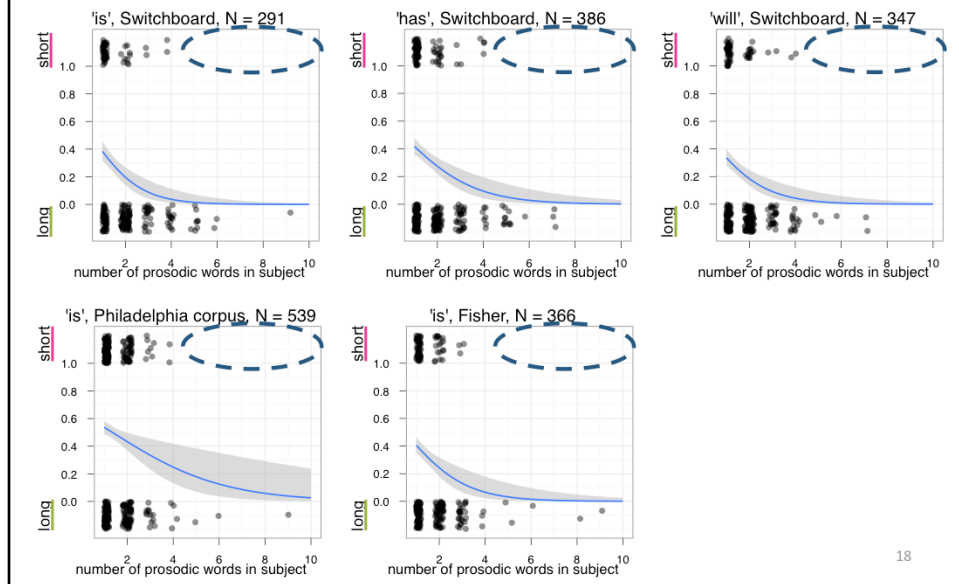
And here's a side-by-side comparison of the plots I just showed you, which drives home that they all show the same downward trend: regardless of which auxiliary is under study, and regardless of which corpus the data is taken from, contraction is less likely after long subjects.

But not only that: in addition to this tapering-off effect, there actually appears to be an apparent cut-off effect, by which, in each case, we find no instances of contraction after subjects of longer than 8 words, and that's what these blue ovals are showing you.

Now, orthographic words is just one possible measure of subject length. I've actually, along with my collaborator Constantine Lignos, tested a number of other measures, which we'll be talking about at the Georgetown Round Table on Linguistics, but to give you a preview, using a mixed-effects model, we don't find a significant effect of number of syntactic phrases or subject duration on contraction, once the effect of orthographic words has been factored out of those measures.

However, there is a significant effect of one other measure, which is number of prosodic words. I counted this by following Selkirk 1995 and took prosodic words to be just all non-function words, so strictly speaking, this is just a quick and dirty measure of non-function words in a subject. The graphs looked like this:

Subject length: prosodic words



The same downward trend is again in evidence, and this time the apparent cut-off is at 4 words.

So, again, we find significant effects of both orthographic and prosodic words on contraction, and this is even after having residualized one on the other, to counteract any effects of correlation between the predictors.

So this is effect of subject length is clearly really strong, and it raises questions about how it's incorporated into speakers' mental representations of this alternation between short and long allomorphs.

In order to start addressing this question, we can take a look at how previous researchers have modeled the factors they found to condition contraction.

Incorporating conditioning factors

Labov 1969

$$\begin{bmatrix} +\text{voc} \\ -\text{str} \\ +\text{cen} \end{bmatrix} \rightarrow (\emptyset) / \begin{bmatrix} * \text{pro} \\ \alpha \text{V} \end{bmatrix} \# \begin{bmatrix} - \\ +\text{T} \end{bmatrix} C_0^1 \begin{bmatrix} \alpha \text{Vb} \\ \beta \text{gn} \\ -\gamma \text{NP} \end{bmatrix} \begin{matrix} \downarrow \\ \uparrow \end{matrix} \leftarrow$$

“enlargement of the concept ‘rule of grammar’” (737)

19

Specifically, I’m thinking of Labov 1969, which is the earliest quantitative work on this variable.

Labov analyzed contraction as a phonological rule of vowel deletion, which was the most common analysis of this variable until the 80s, after which people started treating it as allomorphy, the analysis I’ve adopted here. But it’s still informative to see how Labov went about it.

He modeled it with the variable rule that you see here, and those factors that he found to account for whether or not contraction applied were incorporated as conditions on that variable rule: so, whether the auxiliary’s subject is a pronoun or a noun phrase; whether a vowel precedes the auxiliary, and what type of complement follows the auxiliary.

In this paper, Labov describes variable rules like this one as an “enlargement of the concept ‘rule of grammar’”. Crucially, and this is a point also made by Cedergren and Sankoff in their 1974 paper, variable processes like this are to be treated as part of speakers’ competence, and must be represented in speakers’ grammars alongside categorical rules.

The question I’m going to pursue in the rest of this talk is whether this is an appropriate way of treating the subject length effect on contraction, and, if not, how it should be treated instead.

Now this thread initiated by Labov was taken up again more than twenty years later by Guy and Boberg in their 1997 paper, where they made a similar argument that

“The same things”

- Preceding segment effect on t/d-deletion

/nt/ > **/st/** = **/pt/** > **/ft/** > **/lt/** (Guy & Boberg 1997)

an OCP effect: more features shared with preceding segment = more deletion

- cf. categorical stop deletion in Catalan

deletion only in homorganic clusters (Wheeler 2005)

/kɑmp/ [kɑm] ‘field’, **/pɔnt/** [pɔn] ‘bridge’

but **/serp/** [serp] ‘snake’, **/ark/** [ark] ‘arch’

20

Now, what are those “same things”? Guy and Boberg specifically discuss the variable of t/d-deletion in English word-final consonant clusters, which shows an effect of preceding segment, as follows: the most deletion is found after /n/, then sibilants and stops, then non-sibilant fricatives and finally /l/.

They analyze this as an effect of the Obligatory Contour Principle: the more features a /t/ or /d/ shares with the segment that precedes it, the more likely that alveolar stop is to delete. So /n/, which shares [+coronal] and [–continuant] with /t/, is a strong promoter of deletion, while /l/, which shares only [+coronal], is only a weak promoter.

Guy & Boberg's point is that this effect has clear analogs in attested categorical phenomena. They don't happen to provide any specific examples, but we can find one in stop deletion in Catalan, an invariable process, in which deletion occurs only in homorganic clusters.

So the Catalan example can be taken as a sort of categorical version of variable t/d-deletion: in each case, a process of deletion is conditioned by featural similarity between the deleting stop and the segment that precedes. The fact that the same factor conditions both a variable process like t/d-deletion and a categorical process like Catalan stop deletion is what leads Guy and Boberg to want to say that variable phenomena and categorical phenomena like this should be treated as the purview of the same system.

Sticking with our t/d-deletion example for a moment longer, it's also well-attested that

“The same things”

- Following segment

/t#C/ > **/t#V/** (Guy 1980)

– cf. categorical stop deletion in Catalan (Wheeler 2005)

/ark **b**ajj/ [ar βajj] ‘low arch’

but /ark **a**mples/ [ark ample] ‘wide arch’

/serp **g**ran/ [ser γran] ‘large snake’

but /serp **a**dʒil/ [serp adʒil] ‘quick snake’

21

the segment following the t or d also has a strong effect on that variable deletion, with more deletion before consonants than before vowels.

Although Guy & Boberg don’t discuss it in their paper, this is another example of a conditioning factor that has an analog in categorical alternations, and we can again find examples from Catalan: stop deletion occurs categorically when a consonant follows, but never when a vowel does.

So again, there seems to be good reason for saying that a factor like following segment is encoded in the grammar: this is the kind of environment which we find categorical processes responsive to, also. In both of these t/d-deletion cases, a variable phenomenon appears to be conditioned by the same things that condition categorical phenomena.

But getting back to contraction, there seems to be a clear difference between preceding or following segment, on the one hand, and number of words, which we found affecting contraction, on the other. When we think about the subject length effect on contraction, there’s a sense in which the grammar seems to be counting: contraction responsive to precisely how many words are in that subject, to the point that short allomorphs become apparently nonexistent when a subject hits 8 words in length.

–The same things

- Number of (prosodic) words

22

But this is not the same sort of thing that conditions categorical alternations. We don't find, say, phonological processes, or invariable allomorphic alternations, that can count up to 8. So where conditions on variable phenomena are concerned,

This is leading us to...

- Grammar
 - Factors that can condition categorical alternations
- Use
 - Factors that can't condition categorical alternations
 - Performance effects: e.g. memory, articulation
(Labov 1969)

23

this seems to be leading us toward a split between grammar and something else, which, for now, I'm calling a system of use.

I'm following Guy & Boberg in treating those factors that are also found to condition categorical alternations, like preceding or following segment, as the purview of the grammar, where variable and categorical alternations are both represented.

But factors like the subject length effect on contraction, which aren't the sort of thing that can condition categorical alternations, I'm attributing to this extragrammatical 'use' component. And I'm following Labov in proposing that the sort of things that are outside the grammar are performance effects, like constraints on memory, or articulation – these are the sorts of things which he described as being “outside the linguistic system proper.”

Now, is it reasonable for the subject length on contraction to be interpreted as a performance effect like this? I think it is. Specifically, I'm pursuing the hypothesis that the subject length effect could be the result of

Production planning

- Ferreira 1991
 - Longer subject = longer initiation time
 - Longer object = no effect on initiation time
- Wheeldon & Lahiri 1997
 - More prosodic words = longer initiation time

24

constraints on production planning.

For instance, Ferreira 1991, in a series of experiments, examined how long it took speakers to repeat back sentences they had memorized. She found that the longer the subject of a sentence, the more time it took before speaker initiated their repetition, implying that they needed to plan that subject before they spoke. But the length of the OBJECT of that sentence did not affect initiation time, which Ferreira took to mean that planning for what followed the subject occurred separately from planning for the subject itself.

This could provide a plausible account for the subject length effect on contraction: if a subject is particularly long, it's conceivable that speakers will have planned only that subject and not the verb that follows it. Now contraction requires a host and its auxiliary to be very close, because it results in cliticization of that auxiliary to its host. So it's conceivable that if both aren't planned at the same time, we can't get the closeness that necessitates that short allomorph insertion.

And our finding that prosodic words are playing a role in contraction also jibes with work such as Wheeldon & Lahiri 1997, who find that, even for sentences which are matched for syntactic complexity, the more prosodic words in a sentence, the longer it takes speakers to initiate it, which they take to indicate that prosodic words are a unit of planning.

There are definitely some interesting directions we can take this in future work to really probe whether it's planning that's having an effect here. For instance, we can

Conclusion

- Effect of subject length
 - Short allomorphs disfavored after long subjects
 - Effect of both orthographic and prosodic words
 - Three auxiliaries, three corpora
- Not what's found to condition categorical alternations
- An extra-grammatical effect, potentially due to production planning

25

by summarizing.

I've demonstrated an effect of subject length on contraction, by which an auxiliary's short allomorph becomes less and less likely to surface as a subject gets longer, with both orthographic and prosodic words being significant predictors of contraction rate, and I've demonstrated this effect for three different auxiliaries in three separate corpora.

Now this subject length effect on contraction is not the sort of thing that can condition categorical alternations. As a result, we may not want to propose that it's encoded in the grammar, unlike what has been done with factors like preceding and following segment on t/d-deletion.

Instead, it may be that the subject length effect is extra-grammatical, potentially the result of a system of production planning.

And the bigger take-home point here is that our models of variation need to be able to account for a range of different types of conditioning factors, and that it may not be principled to treat all factors that we find conditioning variation as the purview of a single system.

Thank you!