

A lexical OCP effect for *that*?

The assigned readings for Lecture 5 all concern the production and processing of the word *that* as complementizer versus determiner. In the earliest study, Tabor *et al.* (1997) find that the preferred interpretation depends upon the position of the item in the utterance. Utterance-initial tokens are preferentially interpreted as determiners, while those following the verb are more often construed as complementizers. This asymmetry was not predicted by processing theories extant at the time, but (according to the authors) follows from differences in the frequency of the two interpretations in the respective structures. Gibson argues that while listeners do keep track of relative probabilities in syntactic expectations, it is not an effect of this tracking that derives the complementizer/determiner positional asymmetry for *that*. Rather, listeners employ their knowledge of context-independent lexical frequency in choosing among possible interpretations.

It would be interesting to see the extent to which sentences with double-*that* sequences occur. That is, the stimulus set could have included sentences such as those in 1c, which include *that* being used as both a complementizer and a determiner consecutively, as well as those in 1a and 1b:

1a. *Determiners*

That cheap hotel was clean and comfortable to our surprise.

The lawyer insisted that cheap hotel was clean and comfortable.

1b. *Complementizers*

That cheap hotels were clean and comfortable surprised us.

The lawyer insisted that cheap hotels were clean and comfortable.

1c. *Doubles*

That that cheap hotel was clean and comfortable surprised us.

The lawyer insisted that that cheap hotel was clean and comfortable.

As far as processing goes, inclusion of two *that* tokens disambiguates the utterance far earlier than would otherwise occur (for these sentences, three words earlier than when the following verb otherwise disambiguates). Thus there is a functional reason that inclusion of a second *that* might be desirable. (The repetition might at times be interpreted first as a speech error, but in an experimental setting this is unlikely since the subjects' real world knowledge would tell them that the stimuli were purposefully and carefully produced.)

On the other hand, it involves extra articulatory effort, which *ceteris paribus* might militate against its inclusion. In addition, one might expect a "lexical OCP effect" to be in operation, which disprefers sequences of identical words. It is my impressionistic observation, however, that people still produce these all the time, as in sequences of

auxiliary/main verb *do* which are quite common (*i.e.* “I know what they don’t do in their secret clubhouse, but I wonder what they *do* do.”)

Such a production study would first need to tabulate clause-initial occurrences of determiner/pronoun *that*. It would then need to see how many of those were preceded by a complementizer *that*. (Judging from the frequency data reported in Gibson *et al.*, virtually none of the sentence initial ones, at least, would be. But the situation might be different sentence-internally). One would then determine the ratio of those including a complementizer token to those that do not. This would then need to be compared to the general frequency of occurrence of complementizer *that* where possible.

Luckily, detailed production information on its occurrence is already available in a study by Florian Jaeger and colleagues (on which he reported at ling-lunch last week). Their study includes not only overall frequency of optional *that*-inclusion (which is roughly 50-50). Their breakdown of syntactic context effects also demonstrates more skewed distributions depending, for example, on the amount of material that follows the clause boundary. It would also be interesting to see if such considerations affect the proportion of double-*that* sequences (assuming enough tokens were available that this would be possible).