



# Multiple *That*: A strategy for reducing integration costs

Laura Staum and Ivan A. Sag (Stanford University)

Contact: lstaum@stanford.edu



## Abstract

Sentences with multiple complementizers like *I told him that for sure that I would come* are universally regarded as ‘ungrammatical’ by grammarians, though they often appear in speech, and even in writing. Do these examples reflect a disfluency? Are they actually grammatical? Or are they motivated by processing difficulty? If the repetition of *that* is a production strategy used to reduce integration costs in the complement clause (CC), it should decrease reading times on the subject of the complement clause compared to sentences with only one *that*. To test this prediction, we conducted a self-paced reading study of **Multiple *That*** sentences. Results showed that when integration costs were high, reading times were faster on the embedded subject in Multiple *That* sentences compared to those with only one *that*, suggesting that the extra *that* helps readers understand hard-to-process CCs. If Multiple *That* is not generated by the grammar, then it is an interesting example of an “acceptable ungrammatical” sentence type (Langendoen and Bever 1973).

## The Question

Can processing difficulty motivate the “acceptable ungrammatical”?

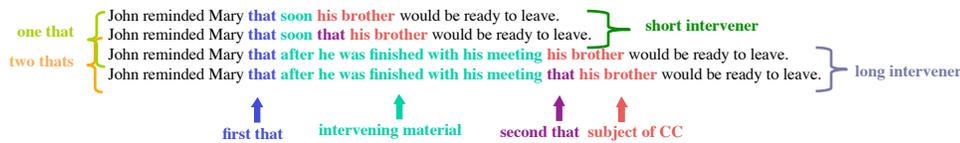
In naturally occurring examples like (1) and (2), a second occurrence of the complementizer *that* appears before the subject of the complement clause:

- (1) They were so cold **that** if they were sitting on the launch pad in this aluminum tank **that** they would form sheets of ice on the outside. (NPR Morning Edition, 7/12/05)
- (2) I truly wish **that** if something like that were to happen **that** my children would do something like that for me. (Switchboard Corpus)

This “extra” complementizer does not provide any new information, sounds unacceptable to many people, and is not an option made available by any grammar (formal, pedagogical, etc.) of English, yet it appears frequently, and it often doesn’t bear any of the phonetic hallmarks of a disfluency (Shriberg 1995). Is it an unusual disfluency? Is it a fluent performance error? Or is it a production strategy for dealing with processing difficulty?

## Design

In a masked, self-paced reading study with 28 participants, we measured reading times on the head noun of the subject of the complement clause (always the second word in the subject NP). Each sentence contained an adverbial between the complementizer and the beginning of the complement clause that was either **short** (one word long) or **long** (seven words long); in addition, each sentence contained **one that** (before the adverbial) or **two thats** (before and after the adverbial).



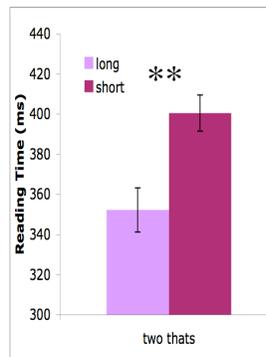
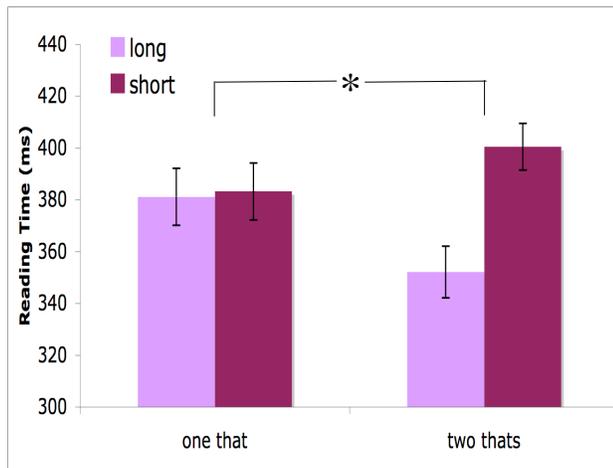
## Predictions

The cost of integrating the subject of a CC increases as the distance between the complementizer *that* and the CC subject increases (following Gibson 2000). Because the extra *that* in Multiple *That* examples is closer to the items to be integrated than the original complementizer, its presence should minimize distance-based integration costs at the subject of the CC by reactivating the dependency link between the verb and its complement. When the adverbial is long, the costs to be reduced are high, and when it is short they are low, making the extra *that* more helpful when the adverbial is long than when it is short. This predicts an interaction between the two experimental factors (presence of an extra *that* and length of the adverbial): If inserting *that* immediately before the subject minimizes distance-based integration costs, then the second *that* should improve reading times on the subject of the CC when an intervening adverbial is long, but not when it is short. Alternatively, if the extra *that* is a fluent performance error that serves no processing function, the second *that* should never improve reading times. Finally, if the extra *that* is merely an unusual disfluency, it should make processing slower after the long adverbials, rather than faster, since a repetition or restart after a longer period signals a larger disturbance.

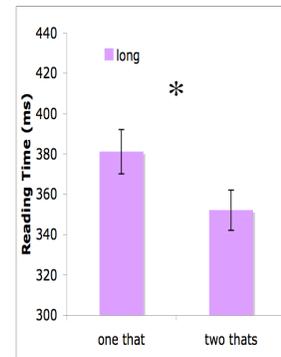
## Results and Discussion

Reading times showed that the subject of the CC was read faster after an extra *that* when the adverbial was long, but not when it was short. This interaction was significant both by subjects ( $F(1,27)=5.60, p=0.025$ ) and by items ( $F(2,1,19)=7.00, p=0.016$ ). A disfluency analysis of Multiple *That* examples cannot account for this difference between the adverbial length conditions. Ungrammaticality could account for a similar interaction: obvious grammaticality violations make structures harder to process, and when the two *thats* are far apart, the violation of the grammar may be less obvious, producing less difficulty. However, ungrammaticality alone can’t predict that the embedded subject in the extra-*that* version should ever be read **faster** than the single-*that* version; an effect of distances between dependents and their links is necessary to account for this result.

Reading time on the subject of the complement clause



When there are two *thats*, reading times on the subject are significantly faster after a long intervener than after a short one ( $t(1,27)=3.3, p=0.001$ ). ⇒ **Longer adverbials are beneficial. This is not predicted by a disfluency analysis.**



When the intervening adverbial is long, the subject of the complement clause is read significantly faster in the two-*that* sentence than the one-*that* version ( $t(1,27)=2.2, p<0.02$ ). ⇒ **The extra *that* actually helps people process difficult clauses! This is not predicted by ungrammaticality alone.**

## Conclusions

Despite violating grammatical constraints, repetition of *that* helps readers process difficult complement clauses.

Processing difficulty can motivate “acceptable ungrammatical” utterances.