The Advantage of the Ungrammatical

Laura Staum Casasanto (lstaum@stanford.edu)
Department of Linguistics, Margaret Jacks Hall
Stanford, CA 94305 USA

Ivan A. Sag (sag@stanford.edu)
Department of Linguistics, Margaret Jacks Hall
Stanford, CA 94305 USA

Abstract

Sentences with multiple complementizers like *I told him that for sure that I would come* often occur in speech and even in writing, although they are not generated by any formal grammar. Here we conducted an acceptability study and a self-paced reading experiment to test whether these 'Multiple *That*' constructions are acceptable, and whether they are motivated by processing difficulty. Results showed that the presence of an extra complementizer always reduced the acceptability of sentences relative to single-complementizer versions, suggesting that this construction is not licensed by the grammar. However, the penalty incurred by the extra complementizer was smaller when more material intervened between the verb and the embedded clause, making integration costs high. In addition, reading times were faster on the embedded subject in Multiple *That* sentences compared to sentences with only one *that* in these more difficult sentences, suggesting that the extra *that* actually helps readers understand hard-to-process complement clauses. Multiple *That* need not be generated by the grammar under a theory of performance that allows processing pressures to add structures to the set of possible sentences.

Keywords: Sentence processing; syntax.

Introduction

The competence/performance distinction (as described in Chomsky and Miller, 1963) has fallen under attack in the face of empirical results demonstrating the complex relationships among acceptability, grammaticality, and processing. Proposals for the total elimination of the competence component (O'Grady, 2007) suggest that characteristics of language previously attributed to grammar can be accounted for by parsimonious constraints on a parser. Even researchers who do not favor doing away with grammar have proposed processing explanations for phenomena previously understood in grammatical terms (Arregui, Clifton, Frazier and Moulton, 2006; Hofmeister et al. 2007). These empirical results suggest reevaluating the interaction between competence and performance in the explanation of linguistic data.

In the current study we investigated a new phenomenon, which we call Multiple *That*, in which speakers produce sentences that contain more than one complementizer to introduce a single complement clause.

In natural speech, the complementizer *that* generally appears between a complement-taking verb like *agreed* in (1) below, and the subject of the complement clause like *he* in (1).

> (1) I agreed **that** he was too late.

Formal grammars generate complementizers exclusively at the beginning of complement clauses. While there is variation in whether or not an overt complementizer appears in this position, grammatical constraints require that when there is an overt complementizer, it must appear immediately before any material that is part of the embedded clause. Although this is the only restriction that is used to specify the location of the complementizer, in this type of sentence, material that intervenes between the verb and the complement is not all that common. Because of this, the complementizer is usually contiguous with both the verb that subcategorizes for it and the subject of the clause it introduces, and by virtue of this contiguity, it can serve as a signal of two things.

First, it clarifies the argument structure of the verb by signaling that a clause is upcoming and that this clause will serve as the complement of the verb; it usually appears in exactly the position that a noun phrase (NP) would appear if the verb had an NP complement:

> (2) I told him {a story, that I was coming}.

Many of the verbs that can take a sentential complement as an argument can also take an NP complement, and so the argument structure of the verb is ambiguous until there is some clear sign that the material following the verb is part of either a clause or an NP. An overt complementizer in speech constitutes such a clear sign, and allows the listener to assign an argument structure to the verb immediately upon interpreting it.

Second, it signals that the subject of the clause is immediately upcoming, making this subject highly predictable and easier to process when it appears. An efficient parser could learn, by tracking statistics, to interpret the complementizer as a signal of these two things, thereby making it easier to assign an argument structure to the matrix verb and to integrate the subject of the complement clause into the parse when it appears.

Importantly, it is only possible for a single item to perform this dual function when the verbal complement

position (immediately after the verb or another object of the verb) and the position immediately before the subject of the clause are the same. If there is material, such as an adverbial, intervening between the verb and the subject of the embedded clause, it is impossible for a single complementizer to serve both of these functions maximally well.

In this circumstance, presumably the grammar still produces only one complementizer, since the grammatical constraints on complementizer generation are not affected by intervening material. If the adverbial modifies the matrix clause, the complementizer will be generated after the adverbial:

(3) I agreed with all of my heart **that** he was too late.

If the adverbial modifies the embedded clause, the complementizer will be generated before the adverbial:

(4) I agreed **that** although he had attempted to introduce me to the woman he was too late.

These solutions both satisfy the grammatical constraint that the complementizer should appear at the beginning of the complement clause. They fail, however, to simultaneously satisfy both of the proposed processing constraints that the complementizer should appear immediately after the verb to signal its argument structure, and that it should appear immediately before the embedded subject to signal its appearance and aid in its integration.

The complementizer in (3) fails to signal the argument structure of the verb right away, leaving the listener to wonder during the adverbial whether there will ultimately be a finite complement clause or some other kind of complement such as *to meet his demands* or *with his assessment*.

Similarly, the complementizer in (4) fails to signal the appearance of the subject of the complement clause, leaving the listener unsure whether the adverbial will continue further at the point when the subject appears.

While in both of these cases the appearance of the subject of the complement clause itself generally disambiguates between the possible outcomes, having to do the work of disambiguating syntactic structures at the same time as integrating the information contained in the subject violates the principle of Uniform Information Density (Levy and Jaeger, 2007) – it's more difficult for the listener to accomplish these tasks all at once than it is for her to do the syntactic disambiguation when she encounters the complementizer and then do the semantic and syntactic integration when she encounters the subject itself.

Thus although sentences like (3) and (4) above do not violate any grammatical constraints, they do violate processing constraints, leaving speakers and listeners with a problem to solve – when there is intervening material between the complement-taking verb and the subject of the complement, they are stuck between a grammatical rock and a processing hard place.

So, do speakers always conform to grammatical constraints in these circumstances? Although the grammar produces only one complementizer regardless of what material might intervene between the verb and the embedded subject, real speakers sometimes produce two complementizers, especially when this intervening material is long. In naturally occurring examples like (4) and (5), a second, 'extra' occurrence of the complementizer *that* appears before the subject of the complement clause:

(5) 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)

(6) 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 is not an option made available by any principled grammar (formal, pedagogical, etc.) of English, and in fact it does not provide any new grammatical information – it simply reiterates information provided by the original complementizer. Yet it appears frequently in speech and even in writing, and it doesn't bear any of the phonetic hallmarks of a disfluency (Shriberg 1995). In addition, the fact that it seems to appear more frequently when the intervening material is long bears explaining. Does the grammar need to be modified to generate this structure, and if so, how could it generate it preferentially in non-local situations? Alternatively, can it be accounted for as a production strategy for reducing integration costs in the complement clause, in a way that naturally accounts for potential locality effects?

In Experiment 1, we'll investigate whether the extra *that* reduces the acceptability of sentences relative to versions with a single complementizer, to evaluate our proposed grammatical constraint against multiple complementizers introducing the same complement clause, and to see if the locality difference in distribution appears in acceptability judgments. In Experiment 2, we'll look for evidence of our proposed processing constraints, to find out whether speakers might have a motivation to flout the grammatical constraint, producing the multiple complementizers we observe in spontaneous speech and writing.

Experiment 1: Judgments

Experiment 1 explores the acceptability of Multiple *That* sentences. First, does an extra complementizer reduce the acceptability of a sentence, indicating that it violates a grammatical constraint? And if so, is this acceptability penalty modulated by the locality of the violation (i.e. the distance between the two complementizers)?

Methods

Participants Thirty native English-speaking Stanford

University students participated in exchange for course credit or payment.

Materials and Procedures A list of twenty items consisting of main clauses containing complement-taking verbs was constructed. 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 either one *that* (before the adverbial) or two *that*s (one before and the other after the adverbial):

(7a) John reminded Mary that soon his brother would be ready to leave.
(7b) John reminded Mary that soon that his brother would be ready to leave.
(7c) John reminded Mary that after he was finished with his meeting his brother would be ready to leave.
(7d) John reminded Mary that after he was finished with his meeting that his brother would be ready to leave.

Sentences were presented one word at a time in a masked, self-paced display on a computer screen using the experimental software package Linger. Participants rated the acceptability of these 20 Multiple *That* sentences, along with 52 filler sentences, on a scale from one (totally acceptable) to seven (totally unacceptable).
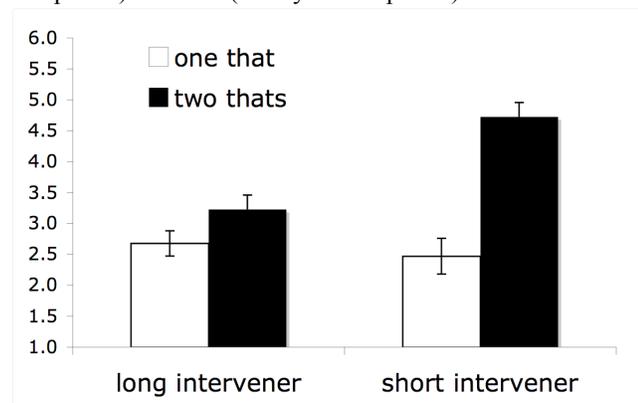

Figure 1: Experiment 1 results, acceptability ratings (scale 1-7).

Results and Discussion

The non-local violations were more acceptable than the local ones, and this difference was significant by both subjects (t1(1,29)=4.58, p=.00004) and items (t2(1,19)=4.43, p=.0001). There was no difference in the corresponding one-*that* conditions in either subjects (t1(1,29)=.96, ns) or items (t2(1,19)=.71, ns), yielding a significant interaction between locality and number of complementizers (F1(1,28)=16.78, p=.0002, F2(1,18)=17.58, p=.0002; Fig. 1). However, even in the non-local condition, the sentences with an extra complementizer were less acceptable than those with a single complementizer (t1(1,29)=2.77, p=.004, t2(1,19)=2.18, p=.02), indicating that this construction constitutes a violation of some grammatical constraint of English.

Because the local and non-local extra *that*s equally violate our proposed grammatical constraint, this influence of locality on the acceptability of Multiple *That* suggests that perhaps the more local violations are more noticeable than those where the two complementizers are farther apart.

While this is a very plausible effect, another possible explanation for this pattern exists: if, as suggested above, the extra complementizer serves a function, its effectiveness at serving this function may be modulated by locality. For example, if the extra *that* serves the function of reactivating representations that were activated by the first *that*, these representations will have decayed more the farther apart the two complementizers are. Thus, the greater the distance between the two *that*s, the greater the functional utility of the second that should be. If this is the case, the less local cases of Multiple That might be more acceptable than the more local ones because they are more functionally justified – that is, they are more helpful.

These two possible explanations for the locality effect, one driven by grammatical constraints and the other driven by functional constraints, make the same predictions with respect to the interaction observed in the acceptability judgments. However, the functional explanation makes an additional prediction, one which is not made by the grammatical explanation, about the influence an extra complementizer will have on processing (relative to single-*that* cases). If locality has an effect solely because it reduces the severity/noticeability of a grammatical violation, then it should be impossible to find a situation in which the extra complementizer ever aids processing – a penalty for grammatical violation could be reduced to nothing, but it could not be negative, so the extra *that* sentences could at best be indistinguishable from the corresponding single *that* versions, or they could be harder to read. Conversely, if locality has an influence at least in part because it modulates the functional utility of the extra complementizer, it should be possible to observe situations in which the extra complementizer helps more than it hurts, and the extra *that* sentences could actually be easier to read than the corresponding single *that* versions.

Experiment 2: Reading Times

Experiment 2 investigates the processing consequences of having an extra complementizer. Does the extra complementizer, which causes a decrement in acceptability, also slow down reading times? Is the influence of extra complementizers dependent on the distance between them? And is it possible to observe a situation in which the extra complementizer provides a benefit in processing, supporting the existence of a functional role for it?

Methods
Participants Twenty-eight native English-speaking Stanford University students participated in exchange for

course credit.

Materials and Procedures Stimuli from Experiment 1 were used in a self-paced reading experiment. Participants read sentences one word at a time in a moving window display and answered a comprehension question about each sentence. Materials were displayed and responses were collected using the experimental software package Linger. We measured reading times on the head noun of the subject of the complement clause (always the second word in the subject NP). Reading times of >1000ms were removed (<1% of all observations).

Results and Discussion

The subjects of the complement clauses were read faster after an extra *that* than after a single *that* when the adverbial was long ($t1(27)=2.22$, $p=0.018$, $t2(19)=2.89$, $p=0.005$), but they were read non-significantly slower after an extra *that* when the adverbial was short ($t1(27)=-0.72$, $p=0.240$, $t2(19)=-1.31$, $p=0.103$). This interaction was significant both by subjects ($F1(1,27)=5.60$, $p=0.025$, Fig. 2) and by items ($F2(1,19)=7.00$, $p=0.016$).

Extra *that* provided a significant benefit in processing the subject when difficulty was high, but it provided no benefit and if anything caused a slight penalty when sentences were already easy to process.
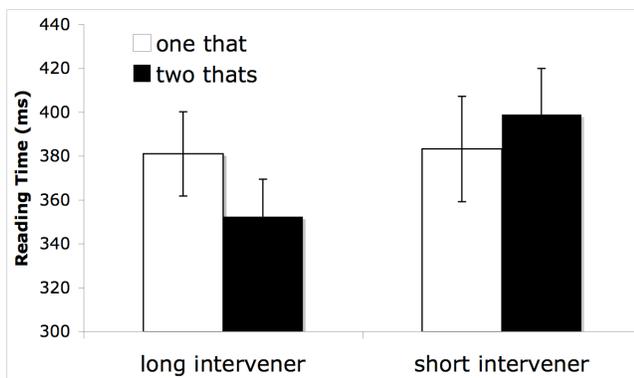
Figure 2: Experiment 2 results, reading times in milliseconds.

As predicted, extra *that* was more beneficial when the intervening material between the verb and the subject of the embedded clause was long than when it was short. Not only was there no observable benefit of extra *that* with the short interveners, but there was a trend in the opposite direction, suggesting that in these cases, if the extra *that* did provide any benefit, it was outweighed by the penalty caused by the grammaticality violation (consistent with the results of Experiment 1). However, it's important to note that this does not entail that there was no benefit of the extra *that* in these circumstances, only that whatever benefit there might have been was not large enough to be observable.

The interaction between the length of the adverbial and the presence/absence of an extra complementizer itself could be accounted for without appealing to production strategies. Perhaps the distance between two items producing a grammaticality violation can modulate the processing consequences of the violation. Presumably, obvious grammaticality violations make structures harder to process, and when the two *that*s are far apart, the violation of the grammar may be less obvious or less memorable, producing less difficulty. If so, then perhaps the extra *that* sentences with long adverbials suffer less of a penalty from the violation of the grammatical constraints, as suggested above. The current experiments have not ruled out such an effect, and it could be operating in tandem with the processing pressures that also predicted this interaction.

However, modulating the penalty for grammatical violations based on distance can't predict that the embedded subject in the extra-*that* version should ever be read *faster* than the single-*that* version; this result demands an account under which extra *that* is beneficial, strongly supporting a processing motivation for the production of the extra *that*.

We propose that the extra *that* is a production strategy speakers use to simultaneously satisfy the processing constraints described above when there is material intervening between the verb and the subject. In the experiments presented here, we considered examples (like 5 and 6 above) in which the intervening material is part of the embedded clause, which are more common in naturally occurring speech and writing. In these cases, if only one complementizer is produced, it appears immediately after the verb, but it is not contiguous with the embedded subject. The parser should therefore be able to use it as a signal of the verb's argument structure, but not as a signal that the subject of the complement clause is immediately upcoming. Thus, in these examples we expect difficulty to arise at the point of integrating the subject of the complement clause, where reading times were measured in this experiment.

If the complementizer's ability to serve as a signal of the embedded subject is modulated by the distance between the two items (the complementizer and the subject), then the cost of integrating the subject should increase as the distance between it and the complementizer *that* increases. A short adverbial will cause a small amount of difficulty, and a long adverbial will cause a larger amount of difficulty. Inserting an extra *that* immediately before the subject of the complement clause should solve this problem in both cases by reducing this distance (and thus the costs) back to zero.

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 exactly the interaction we observed between the presence of an extra *that* and the length of the adverbial: If inserting *that* immediately before the subject of the complement clause minimizes distance-based integration costs (following Gibson 2000), then the second *that* should improve reading times on the subject of the complement clause more when an intervening adverbial is long than when it is short.

These results are highly consistent with corpus and behavioral research about the circumstances in which one

finds optional single complementizers. An overt complementizer (as opposed to a null complementizer) is more likely when a complement clause is less predictable, given both the material that has come before the complement clause and the material at the beginning of the clause (Jaeger, Snider, Staum and Jurafsky, 2006).

It is easiest for a listener to tell that a complement clause is beginning when it begins immediately with its subject; the ease or difficulty of knowing when the complement clause has begun is sometimes quantified as the predictability of the syntactic structure given the lexical item the listener has just encountered. Because this syntactic predictability is high when the clause begins with its subject and low when it begins with something else, the likelihood of an overt single complementizer increases when there is material (such as an adverbial) that intervenes between the complement-taking verb and the subject of the complement clause. The longer this material is, the more likely a complementizer is (Jaeger, 2006). In addition, there is evidence that when the relative complementizer *that* is present, it makes difficult relative clauses easier to process (Jaeger, 2007). We suggest that the processing benefit that readers get from the overt complementizer in single-*that* examples is the same as the benefit provided by the second *that* in Multiple *That* examples.

General Discussion and Conclusions

Traditionally, acceptability has been taken as evidence of grammaticality. Yet studies suggest that some constructions may be unacceptable due to processing difficulty rather than ungrammaticality. The grammar can freely generate structures that don't occur in speech or writing, and even structures that are judged to be unacceptable, so long as they can be ruled out via constraints on the parser. Such parsing limitations can, by virtue of constraints on the parser, eliminate the need for grammatical constraints to rule out these unacceptable utterances (Hofmeister et al., In Press).

In addition to generating unacceptable utterances, a competence grammar might also fail to generate utterances that do occur. Such "acceptable ungrammatical" utterances (Langendoen and Bever, 1973) arise when they are sufficiently easier than a fully grammatical alternative from a parsing perspective. They may even achieve full acceptability, as Langendoen and Bever argue is the case for utterances containing `a not unhappy man'.

The arena of verb-phrase (VP) ellipsis has served as another test case of this theory of grammatical undergeneration. While categorical syntactic accounts of VP ellipsis notoriously undergenerate relative to the attested and acceptable examples, ease of constructing an appropriate antecedent for a verb phrase ellipsis does a good job of accounting for the acceptability of examples that lack matching antecedents (Arregui et al., 2006). Examples that are easy to analyze in syntactic terms are more likely to be acceptable according to elicited judgments.

This phenomenon is a good example of how an explanatory account might be achieved via the interaction of a categorical syntactic grammar that undergenerates and a theory of strategies that the processor employs along with the grammar and other resources. However, if one defines ease of processing in syntactic terms, these data do not necessarily argue against a competence-based account where gradient ungrammaticality stems from gradient grammatical constraints.

Our results suggest that although it adds no new information, the extra *that* in Multiple *That* examples is an effective strategy for reducing integration costs in the complement clause when they are high (the predictive power of the complementizer can outweigh the penalty from a non-local grammaticality violation in such an interactive system). In addition, they suggest strongly that both grammatical and processing constraints are at work in determining whether one *that* or two will be produced. Although we have not ruled out the possibility that extra *thats* could be generated by a competence grammar, the fact that they produce a decrement in acceptability judgments even when they are aiding processing provides a serious challenge for theories in which performance factors account for all linguistic behavior without input from a competence grammar.

The results presented in this paper are thus most consistent with a linguistic system in which grammatical and processing constraints interact with one another to produce the linguistic behavior that we observe. Multiple That doesn't need to be generated by the grammar if we take a view of the competence-performance relation that cuts both ways: Not only can processing constraints diminish the acceptability of fully grammatical sentences, but they can also promote the acceptability of ungrammatical sentences.

Acknowledgments

References

Arregui, A., Clifton, C., Frazier, L., and Moulton, K. (2006). Processing elided verb phrases with flawed antecedents: The recycling hypothesis. *Journal of Memory and Language* (pp. 232-246). Elsevier.

Chomsky, N. and Miller, G. A. (1963). Introduction to the formal analysis of natural languages. In Luce, R. D., Bush, R. R., and Galanter, E. (Eds). *Handbook of Mathematical Psychology, Vol. 2*, 269-321. New York: Wiley.

Gibson, E. (2000). The dependency locality theory: A distance-based theory of linguistic complexity. In Miyashita, Y., Marantz, A., and O'Neil, W. (Eds.), *Image, language, brain*, 95-126. Cambridge, MA: MIT Press.

Hofmeister, P., Jaeger, T. F., Sag, I. A., Arnon, I., and Snider, N. (In Press). Locality and Accessibility in Wh-questions. In Featherston, S. and Sternefeld, W. (Eds)

Roots: Linguistics in Search of its Evidential Base. Berlin: de Gruyter.

Jaeger, T. F., Snider, N. Staum, L. and Jurafsky, D. (2006). (In)dependence of Lexical and Syntactic Production: *that*-reduction and omission in spontaneous speech. Presented at CUNY 19, New York, NY.

Jaeger, T. F. (2006). *Redundancy and Syntactic Reduction in Spontaneous Speech*. Doctoral Dissertation, Department of Linguistics, Stanford University.

Jaeger, T.F. (2007). Rational Speakers: speakers help processing when it is most necessary. Presented at the 13th Architecture and Mechanisms for Language Processing 2006, Turku, Finland.

Langendoen, D. T. and Bever, T. G. (1973). Can a not unhappy man be called a not sad one? In S.R. Anderson & P. Kiparsky (Eds.), *A festschrift for Morris Halle,* (392-409). New York: Holt, Rinehart and Winston, Inc.

Levy, R. and Jaeger, T. F. (2007). Speakers optimize information density through syntactic reduction. *Advances in neural information processing systems (NIPS), 19*. 849-856. Cambridge, MA: MIT Press.

O'Grady, W. (In Press). The emergentist program. *Lingua*. Elsevier.

Shriberg, E. E. (1995). Acoustic properties of disfluent repetitions. *Proceedings of the International Congress of Phonetic Sciences, 4*. 384-387. Stockholm, Sweden.