



Optional *that* indicates production difficulty: Evidence from disfluencies

T. Florian Jaeger

Stanford University, USA

Abstract

Optional word omission, such as that omission in complement and relative clauses, has been argued to be driven by production pressure (rather than by comprehension). One particularly strong production-driven hypothesis states that speakers insert words to buy time to alleviate production difficulties. I present evidence from the distribution of disfluencies in non-subject-extracted relative clauses arguing against this hypothesis. While word omission is driven by production difficulties, speakers may use that as a collateral signal to addressees, informing them of anticipated production difficulties. In that sense, word omission would be subject to audience design (i.e. catering to addressees' needs).

1. Introduction

Optional word omission, as optional that omission in complement clauses (1a) and non-subject-extracted relative clauses (1b), has been one of the testing grounds for audience design (i.e. the questions *when*, *how* and *to what extent* speakers react to addressees' needs [3]).

- (1a) I believe (that) my brother stole your bike.
(1b) I mean everything (that) they spray out in the fields.

In earlier work with Thomas Wasow on non-subject extracted relative clauses (henceforth NSRCs) [17], we have shown that that omission correlates with the distribution of disfluencies. The current project compares two competing accounts of that correlation. This comparison pertains to the question of the extent to which audience design influences word omission. Do speakers mention that to *alleviate* production difficulties (as first suggested in [21])? Or, alternatively, does that serve as *collateral signal* [5] to the hearer that the speaker anticipates production difficulties? Before I describe these two hypotheses in more detail, I briefly summarize the relevant discussion in the literature.

Research on the influence of audience design on word omission has focused on the facilitation of comprehension, most specifically the avoidance of structural ambiguity [2, 25]. Consider (1a). If that is omitted, it creates temporary structural ambiguity (the NP my brother could also be an object argument of believe, as in “I believe my brother”). Indeed, reduced complement clauses take longer to read than their full-form counterparts [22] (I use the terms “full” and “reduced” to refer to complement and relative clauses with and without that, respectively). For this reason, using the full form has been argued to serve the facilitation of comprehension [2, 11-13, 25] (which would make word omission subject to audience design). However, evidence from several laboratory production studies [7, 14, 18, 21] and corpus studies [17, 21, 24] argues against such comprehension-facilitation accounts of word omission. Ferreira & Dell [7] found no evidence that speakers avoid reduced forms when this leads to structural ambiguity (not all reduced forms result in ambiguity). While this finding is a null effect and as such should be interpreted with caution, it has been replicated [18, 24].

Ferreira & Dell [7] propose that word omission is subject to availability-based sentence production: speakers use the reduced form when the material following the omitted word is readily available. This hypothesis has received a fair amount of support. For example, speakers are more likely to use full-form complement/relative clauses when the embedded subject is more complex (e.g. a full lexical NP rather than a pronoun) and therefore takes longer to plan [8, 16, 21, 24]. Elaborating on the idea of availability-based sentence production, Race & MacDonald [21] hypothesized that speakers insert that to *alleviate* production difficulties. Uttering a relativizer may give speakers more time to overcome production problems in the embedded clause. I dub this the Alleviation Hypothesis.

Alternatively, mentioning that (where it may be omitted) could be a *collateral signal* (as defined in [5]). More specifically, speakers may use that to signal that they anticipate production difficulties. I dub this the Signal Hypothesis. The Signal hypothesis is based on and informed by earlier research on collateral signals of production difficulty [6, 9, 10]. Consider the case of uh/um. Speakers intentionally use the fillers uh/um when they are likely to suspend speech [6], and addressees are sensitive to this signal [3, 6]. This means, even though the distribution of uh/um is driven by production difficulties, the use of uh/um is a case of *audience design*. The Signal Hypothesis applies this insight to optional word omission. Consistent with existing evidence, the Signal Hypothesis predicts that the distribution of that is driven by production pressures. But, unlike the Alleviation Hypothesis, the Signal Hypothesis attributes this correlation to audience design: mentioning that is a collateral signal to addressees.

The two hypotheses make different predictions about the distribution of disfluencies (as direct evidence of production difficulties) in full and reduced clauses. The Alleviation Hypothesis predicts that mentioning that at the beginning of a complement or relative clause reduces disfluency in the clause. The Signal Hypothesis makes the opposite prediction. To compare the two hypotheses, I conducted a large-scale corpus study of disfluencies in full and reduced NSRCs (future research will show whether the observations made here also apply to other word omission environments). While it is beyond the scope of this paper to show whether addressees are sensitive to relativizer presence (a prerequisite of the Signal Hypothesis), I examine whether addressees *could* interpret relativizers as signals of upcoming production difficulties.

Section 2 gives an overview of the database and methodology. Section 3 tests the Alleviation Hypothesis. Section 4 provides a preliminary test of the Signal Hypothesis. Section 5 incorporates the new findings into a model of relativizer omission. The implications of these studies are discussed in Section 6. Directions for future work are addressed in Section 7.

2. Method and data overview

The results presented here are part of ongoing work [15, 27] on all 4,400 NSRCs from Paraphrase version of the Treebank III Switchboard corpus [4]. The corpus consists of 650 parsed and part-of-speech-tagged telephone conversations on selected topics between two strangers. All conversations in the Switch-

board corpus are annotated for disfluencies [20]. For the current study, NSRCs with *wh*-relativizers (*which*, *where*, *who*, etc.) were excluded because tests revealed that omission of these relativizer was frequently unacceptable. Of the remaining 3,701 NSRCs, 1,601 (43.3%) were full NSRCs (i.e. with *that*); 2,100 were reduced NSRCs. All NSRCs and the information analyzed below were automatically extracted using Tgrep2 [23] and Perl scripts (available upon request).

2.1. Data overview

Of the 3,701 NSRCs in the dataset, 593 (16%) contained at least one disfluency (compared to 36% for all types of clauses in the Switchboard corpus). About 1.3% were *part* of a larger disfluency, but this did not affect relativizer omission ($\chi^2 < 1$) and will therefore not be discussed further.

Counting separately each repetition of fillers, and restarts, etc., the NSRCs contained 793 disfluencies. 351 (44%) were fillers, predominantly *uh/um* (61%) and *you know* (32%):

- (2a) ... the nuclear [NSRC *that, uh, they use*] ...
- (2b) ... things [NSRC *that, you know, two people can do*] ...
- (2c) ... every time [NSRC *I, uh, I spent money, I mean, cash*] ...

The remaining 442 disfluencies were restarts (including lexical, syntactic, and covert repairs in the sense of Levelt [19]) or complete suspensions. These repairs were on average 1.35 words long (STDEV = 0.96, ranging from 1 to 9 words).

- (3a) ... the way [NSRC *our system, our court system works*] ...
- (3b) ... some aunts [NSRC *that, uh, I, I do*] ...
- (3c) ... all *the*, [NSRC *that we're doing here*] ...

The NSRCs contained on average 5.3 words (STDEV = 3.9; ranging from 2 to 42 words), resulting in an overall disfluency rate (counting fillers, restarts, and suspensions) of 0.04 disfluencies per word (i.e. every 25th word belongs to a disfluency; STDEV = 0.14, ranging from 0 to 3.5).

3. Testing the Alleviation Hypothesis

The purpose of the first study is to test the Alleviation Hypothesis. According to the Alleviation Hypothesis, uttering *that* buys times for the speaker to reduce production difficulties. The Alleviation Hypothesis does not state how often this additional time is sufficient to prevent a disfluency that otherwise would have surfaced (henceforth the success rate). For empirical evaluation, it is important to distinguish versions of the Alleviation Hypothesis based on the assumed success rate. The strongest Alleviation Hypothesis assumes a success rate of 100%. Hence full NSRCs should contain significantly less disfluency than reduced NSRCs. No direct predictions are made for fillers, but as suspensions/restarts are often preceded by fillers, full NSRCs should contain fewer fillers. Furthermore, relativizer presence may correlate with a higher likelihood of disfluencies immediately preceding the NSRC because higher workload during the planning of NSRCs may increase the need to buy time. Before I discuss weaker versions of the Alleviation Hypothesis, I test the strongest Alleviation Hypothesis.

I investigated the distribution of fillers, suspensions, and restarts immediately preceding and following the beginning of an NSRC (with or without a relativizer). For the analysis, all fillers form one group (separate tests for the two most frequent types of fillers *you know* and *uh/um* did not reveal significant differences with regard to relativizer omission). Suspensions and restarts form another group.

3.1. Results: Relativizers correlate with disfluencies

The results are summarized in Table 1. The first two rows list the percentage of full and reduced NSRCs that contain at least one disfluency either in the modified NP (i.e. preceding the NSRC) or within the NSRC itself. The last row gives the significance level of Fisher's Exact test. All results were significant. Full NSRCs are at least two times more likely to contain/to be preceded by disfluencies than reduced NSRCs.

Table 1: Percentage of NSRCs preceded by/containing disfluencies

	Fillers		Suspension/Restart	
	In NP	In NSRC	In NP	In NSRC
% of full NSRC	4.7%	8.8%	2.2%	17.5%
% of reduced NSRC	1.9%	4.2%	1.1%	7.9%
Fisher's Exact	p < 0.001	p < 0.001	p < 0.02	p < 0.001

3.2. Intermediate discussion

The observed correlations between disfluency and relativizer presence are predicted by production-driven accounts of word omission [7, 17, 21]. The positive correlation of relativizer presence with production difficulties preceding NSRCs is compatible with the Alleviation Hypothesis. However, the finding that relativizers correlate with a *higher* likelihood of disfluencies in NSRCs rejects the strongest Alleviation Hypothesis.

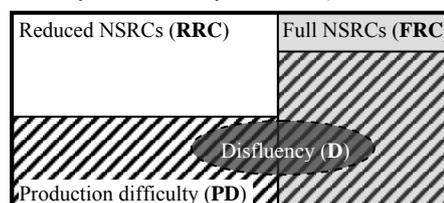
The results presented above are based on pooling all disfluencies in an NSRC. But maybe relativizers only help to alleviate production difficulties that occur at the beginning of NSRCs. In that case, pooling all disfluencies would unfairly bias against the Alleviation Hypothesis. To address this possibility, I conducted separate tests for the presence of suspensions/restarts in the first one to four words of NSRCs. The results, summarized in Table 2, still argue against the strongest Alleviation Hypothesis.

Table 2: Suspensions/restarts within the first 1..4 words of NSRC.

	Word 1	Word 1-2	Word 1-3	Word 1-4
% of full NSRC	5.5%	9.7%	10.9%	12.7%
% of reduced NSRC	2.7%	3.9%	4.9%	5.9%
Fisher's Exact	p < 0.001	p < 0.001	p < 0.001	p < 0.001

Nevertheless, it could be that a weaker version of the Alleviation Hypothesis holds. Assuming a weak Alleviation Hypothesis, where the success rate of alleviation (due to relativizer insertion) is smaller than 100%, the results presented above may be due to an indirect association of *that* with more complex NSRC (e.g. because Alleviation isn't the *only* function of *that*). In that case, the higher likelihood of full NSRCs containing disfluencies could be a side effect. If full NSRCs are on average more complex, they will be associated with more production pressure, and they will more frequently result in production difficulties. To illustrate this idea, consider the Venn diagram in Figure 1 (the diagonal stripes indicate the subset of NSRCs that resulted in production difficulties).

Figure 1: Venn diagram of relativizer distribution, production difficulty, and disfluency in NSRCs (not drawn to scale)



The weak Alleviation Hypothesis predicts that relativizer presence alleviates production difficulties *after other factors are controlled for*. In other words, the proportion of alleviated production difficulties in full NSRCs (i.e. $[[PD \cap FRC] / D]$ divided by $[[PD \cap FRC]]$) should be higher than the proportion of alleviated production difficulties in reduced NSRCs ($[[PD \cap RRC] / D]$ divided by $[[PD \cap RRC]]$). These proportions cannot be directly calculated (since PD is not known), but one can try to control for as many factors as possible that contribute to making an NSRC a member of PD (i.e. factors that make an NSRC complex). Next, I describe the design of the statistical analysis that aims to do that.

3.3. Method and predictions

As in Section 3.1, two separate analyses were performed for fillers and suspensions/restarts. For the current study, a more sophisticated measure than relativizer presence was used as dependent variable: the normalized rate of disfluencies in the NSRC (i.e. the number of disfluent words per word). The number of words that were part of a disfluency was automatically extracted from the Switchboard Corpus. This number was divided by the overall number of words in that domain, yielding a normalized rate of disfluencies. This measure is finer-grained than disfluency presence (used in Section 3.1) and less correlated with NSRC length. Normalized disfluency rates proved a good measure of production difficulty; however, all results were confirmed for other measures of disfluency, cf. Section 3.5. Consider example (3a), repeated below. The NSRC contains six words, two of which are part of a disfluency (underlined). The NSRC's normalized disfluency rate in (3a) is therefore $2 / 6 = 0.33$. The normalized disfluency rate preceding the NSRC in (3a) is $0 / 2 = 0$.

(3a) ... the way our system, our court system works ...

Analyses of variance (ANOVAs) were conducted with subjects as random effects and relativizer as well as control factors as fixed effects. Including speaker effects in the ANOVAs accounts for the fact that speakers may have different base rates of disfluency. Table 3 summarizes the design of the two ANOVAs. The ANOVAs includes factors contributing to the complexity of NSRCs, and factors known to correlate directly with the presence of disfluencies as controls. Next, I briefly describe these controls.

Table 3: Design of ANOVAs

Dependent variable	Normalized disfluency rate in NSRC
Fixed effects [categorical]	Relativizer presence,
	Length of modified NP (in words)
	Is the NSRC subject a pronoun?
[continuous]	Rate of speech before NSRC
	Rate of speech within NSRC
	Length of RC (in words)
	Normalized rates of fillers in NP
	Normalized rate of suspension/restarts in NP
Random effects	Speaker

The rate of speech preceding and within NSRCs was included since rate of speech is known to correlate with disfluencies. The length of NSRCs (in words) was included since more complex NSRCs probably correlate with more disfluencies. Although the dependent variable already includes a control for NSRC length, NSRC length may correlate with a more than linear increase in disfluencies. Next, the normalized disfluency rates preceding the NSRC were included as a factor to control for production difficulties immediately preceding

NSRCs (which may carry over into the NSRC). The NSRC subject's complexity was included because the availability-based sentence production hypothesis ([7], cf. Section 1) predicts that it should have an especially strong effect on relativizer presence. Finally, the grammatical function of the modified NP was included in case the NSRC's position within the matrix clause has an effect on disfluency rates. Other controls were tested but are not included in the final tests since they did not contribute to the model ($F_s > 1$).

If relativizers help to alleviate production difficulty, the normalized disfluency rate for NSRCs with a relativizer should be smaller than the rate for reduced NSRCs.

3.4. Results: Relativizers predict higher disfluency rates

Relativizer presence has a significant main effect on both the rate of fillers in the NSRC ($F(1, 652) = 6.3, p < 0.02$) and the rate of suspensions/restarts in NSRCs ($F(1, 989) = 6.1, p < 0.02$). But, contrary to the Alleviation Hypothesis, relativizer presence (i.e. full NSRCs) correlates with *higher* normalized disfluency rates even after the disfluency rates immediately preceding the NSRC are controlled for (cf. Table 4).

Table 4: Marginal means of the normalized disfluency rates

	Fillers	Suspensions/Restarts
Full NSRC	0.014	0.030
Reduced NSRC	0.005	0.019

NSRC length has a strong effect on both of the normalized disfluency rates ($F_s > 100$). This is due to an exponential increase in the average number of disfluencies for longer NSRCs. Furthermore, the speech rate in the NSRC significantly affects the rate of disfluencies ($F_s > 15$). The effect of the NP's grammatical function approached significance for the normalized rate of suspensions/restarts ($F(3, 2863) = 2.5, p < 0.06$). All other factors failed to reach significance.

3.5. Discussion

The results argue against the Alleviation Hypothesis [21]. Full NSRCs are more likely to contain a suspension/restart (Section 3.1). This finding holds after controlling for other factors that could contribute to the amount of disfluency in NSRCs (i.e., the NSRC's complexity, the production pressure immediately preceding the NSRC, the rate of speech, and speaker effects). Thus even the weak Alleviation Hypothesis is hard to reconcile with the above results. Next, I discuss a couple possible objections to this conclusion.

First, it could be that normalized disfluency rates are not the best way to test the Alleviation Hypothesis. Maybe alleviation would show up in terms of a decrease in the absolute length of disfluencies or a decrease in the absolute number of disfluencies in the NSRC. Additional ANOVAs revealed that this was not the case. Regardless of which disfluency weight measure is chosen as the dependent variable (length, number, normalized disfluency rate), the effect of relativizer presence is either non-existent or in the opposite of the direction predicted by the Alleviation Hypothesis. Also, as shown in Section 3.2, it is not the case that relativizers only alleviate production difficulties immediately following them. This result was confirmed by an ANOVA with the factors from Table 3, but with the rate of suspensions/restarts in only *the first five words* as dependent variable. Contrary to the Alleviation Hypothesis, relativizers still were associated with a *higher* rate of suspensions/restarts ($F(1, 893) = 11.1, p = 0.001$).

Interestingly, the results are consistent with the Signal Hypothesis. The presence of a positive correlation between dis-

fluency rates and relativizers after controlling for other factors suggests that speakers use relativizers as signals to the hearer.

4. Testing the Signal Hypothesis

The next question is whether addressees could potentially use relativizer presence as a signal that the speaker will be likely to run into production difficulties (as predicted by the Signal Hypothesis). In other words, given all information addressees have access to before they hear the beginning of an NSRC, does the presence of a relativizer provide additional information about the rate of disfluencies in the upcoming NSRC?

4.1. Method and predictions

Two separate ANOVAs were conducted for the normalized rates of fillers and suspension/restarts within NSRCs. In contrast to the analysis in Section 3, the factors included in these ANOVAs only contained information available to addressees prior to the beginning of the NSRC (cf. Table 5).

Table 5: Design of ANOVAs

Dependent variable	Normalized disfluency rate in NSRC
Fixed effects [categorical]	Relativizer presence,
	Grammatical function of modified NP
[continuous]	Rate of speech before NSRC
	Length of modified NP (in words)
	Normalized rates of fillers in NP
	Normalized rate of suspension/restarts in NP
Random effects	Speaker

Rate of speech was included as a factor since it usually correlates with disfluency rates. The length of the modified NP was included since producing complex NPs may consume more processing resources, which could cause a processing bottleneck when the speaker plans the NSRC. The grammatical function of the modified NP was included to account for effects that are due to the relative position of the NSRC within the matrix clause. Finally, to capture effects of experienced production difficulty at the point when the speaker is planning the NSRC, the normalized rates of fillers and suspensions/restarts immediately preceding the NSRC were also entered into the ANOVA.

The Signal Hypothesis predicts that relativizer presence exhibits a positive correlation with the rate of disfluencies, even after all information available to addressees prior to the relativizer is controlled for.

4.2. Results: Relativizers predict following disfluencies

Relativizer presence is a significant predictor of a higher rate of fillers ($F(1, 604) = 12.2, p = 0.001$) and a higher rate of suspensions/restarts ($F(1, 946) = 11.2, p = 0.001$).

Only one of the covariates had an effect. Fillers in the modified NP marginally predicted relativizer presence ($F(1, 2942) = 2.8, p = 0.09$). Since fillers often occur in chains, this effect is expected and will not be discussed any further.

4.3. Discussion

The effect of relativizer presence on disfluency rates in NSRCs is predicted by the Signal Hypothesis. The effects are, however, rather subtle. Maybe relativizer presence only informs addressees of disfluencies early in NSRCs? Post-hoc tests were conducted on the rate of suspensions/restarts within the first two/the first five words of the NSRC. As expected, the effect of relativizer presence was stronger (both $F_s = 15.7, P_s < 0.001$). While the results are encouraging for the Signal Hypothesis, there are some issues that deserve discussion.

First of all, if relativizers are collateral signals, addressees should be sensitive to this information. While experiments show readers are sensitive to relativizer presence [11, 21], there is currently no evidence that addressees interpret relativizers as signals of production difficulties. Unfortunately, preliminary data searches only found ten interruptions by the addressee in the middle of an NSRC – too few to test whether interruption is more likely for full NSRCs (as predicted by the Signal Hypothesis). I leave this issue to future research.

Second, two potential confounds to the current study have to be addressed. First, it is well known that speakers lengthen words (including *uh/um*) when they are experiencing production difficulties [6, 9]. Maybe the difference in the relative rate of suspensions/restarts is solely driven by *lengthening* of relativizers (which, of course, can only occur in full NSRCs). Indeed, lengthening of relativizer *that* correlates positively (though only weakly) with the presence of suspensions/restarts at the beginning of an NSRC ($r = 0.09, p < 0.001$). A second potential confound is that the presence of a pause before the NSRC may signal to the hearer that disfluencies are to be expected. However, the existence of pauses is inversely correlated with relativizer presence (only 13% of all full NSRC are preceded by pauses compared to 20% of the reduced NSRC, $\chi^2 = 30.0, p < 0.001$). So, if anything, controlling for pauses should increase the correlation between relativizers and the disfluency rate in NSRCs.

To ascertain that the effect of relativizer presence holds after controlling for the existence of pauses before the NSRC as well as lengthening of *that*, I conducted an ANOVA including these two factors in addition to the factors mentioned above (I am indebted to Neal Snider for extracting information on pauses and the length of *that* from Switchboard). The duration of *that* was transformed by subtracting the mean duration from all cases with a relativizer (this was necessary to avoid collinearity between relativizer presence and relativizer duration). Cases without a relativizer were coded as having a duration of zero (i.e. these cases were treated just like cases with a relativizer of average length).

As expected, both absence of a pause ($F(1, 2854) = 9.4, p < 0.001$) and lengthening ($F(1, 2854) = 20.2, p < 0.001$) are significant predictors of following suspensions/restarts. The effect of relativizer presence was even stronger after controlling for these factors ($F(1, 734) = 20.4, p < 0.001$).

In conclusion, the absence of a pause before an NSRC, lengthening of *that*, and the presence of *that* independently predict that the speaker will run into production difficulties at the beginning of the NSRC.

5. Modeling relativizer omission

The previous study shows that listeners could use the information provided by relativizer presence to predict that the speaker is (more) likely to produce disfluencies in the NSRC. The current study asks whether *speakers'* choice of full over reduced NSRCs is guided by anticipated difficulties and/or the presence of disfluencies preceding the NSRC. In other words, I develop a model of relativizer omission to see whether anticipated difficulty (here estimated by the presence of disfluencies) is a factor that drives the speakers' choice after other factors known to affect relativizer omission are controlled for.

5.1. Method

A Generalized Linear Model predicting relativizer omission was constructed using all factors known to account for a considerable amount of variation in relativizer omission (see A-H below). For more details on A-G, I refer to [8, 15, 16,

27]. H was included since ongoing work (together with Laura Staum) suggests that women use relativizers more frequently than men. Note that the inclusion of F and G assumes that speakers have *some* look-ahead into NSRCs. So does the inclusion of disfluencies in NSRCs as predictors (see below).

- A Grammatical function of modified NP in matrix clause
- B Determiner type of the modified NP
- C Does the modified NP contain a uniqueness-requiring adjective or operator (e.g. the fastest person, ...)?
- D Does a light noun head the modified NP (e.g. way, ...)?
- E Does anything intervene between the modified NP's head noun and the beginning of the NSRC (e.g. my friend from New York that you met ...)?
- F Is the extracted element in the NSRC an adverbial NP?
- G Is the NSRC subject a pronoun (e.g. the friend you met)?
- H Gender of speaker

Earlier logistic regression models of word omission [15, 21, 24] did not model speaker effects. Such models rely on the assumption that each observation in the data set is independent of the other observations. However, datasets often contain several NSRCs from the same speakers, and speakers may have different base rates of relativizer omission (i.e. the assumption of the independence of observations is violated). Indeed, each speaker in the data set on average contributed about 10.8 NSRCs (STDEV = 9.2, ranging from 1 to 44).

I used logit Generalized Linear Mixed Models with A-H as fixed effects and normally distributed random intercepts to model speaker effects (as implemented in the R software library *glmmPQL* [26]; I am grateful to Joan Bresnan for pointing me to this solution). Logit models were used since the dependent variable (relativizer omission) is categorical.

The resulting mixed model (henceforth the standard model) predicts relativizer presence much more accurately (classification accuracy 75%) than the baseline model (a model that always predicts the more frequent event, here relativizer omission; classification accuracy 58%). The improvement was highly significant (change in $-2\log$ -likelihood = 455.0; $p < 0.0001$ based on a χ^2 with DF = 17). All factors contributed significantly to the standard model ($p < 0.01$ for factor C and H; all other factors $p < 0.0001$). Interestingly, the inclusion of speaker effects (STDEV of intercepts = 0.67) considerably improved classification accuracy (from 69% to 75%).

To test whether disfluency presence predicts relativizer presence after all of the above factors are controlled for, a separate model was fit for each of the four disfluency measures (cf. Table 6 below), by adding the disfluency measure to the standard model. Next, the goodness-of-fit of each of these four models was compared against the standard model's fit.

5.2. Results: Disfluencies predict relativizer presence

Table 6 summarizes the change in $-2\log$ -likelihood caused by adding any of the disfluency measures to the standard model as well as the significance level of that change based on a χ^2 with DF = 1 (adding a disfluency measure adds one free parameter to the model).

Table 6: Model improvement for each of the disfluency measures

	Fillers		Suspension/Restart	
	In NP	In NSRC	In NP	In NSRC
Coefficient in model	-0.02	0.89	-0.2	0.55
Change in $-2\log$ -LH	0	19.5	0.4	11.8
Significance level of χ^2	n.s.	$p < 0.001$	n.s.	$p < 0.001$

Information on fillers and suspension/restarts in the NSRC improves the model significantly. As expected, the coefficients for these two significant effects are positive (i.e. the more production difficulty a speaker anticipates, the more likely is a relativizer). Information on disfluencies preceding the NSRC does not improve the model.

5.3. Discussion

Under the reasonable assumption that speakers usually plan the beginning of NSRCs before they start pronouncing the relativizer, they may choose to insert a relativizer if they encounter production difficulties. The above results argue that speakers do this. Even after controlling for other factors known to favor relativizers, fillers and suspensions/restarts in NSRCs are significant predictors of relativizer presence.

Disfluency immediately before NSRCs does not seem to influence relativizer omission. This lack of an effect could be due to insufficient power. Note that there were fewer disfluent NPs (immediately preceding the NSRCs) than disfluent NSRCs. This asymmetry is not surprising since the NPs on average are much shorter (MEAN = 2.3 words, STDEV = 1.1, ranging from 1 to 8 words) than the NSRCs (MEAN = 5.3 words, STDEV = 3.9, ranging from 2 to 42 words). If, however, confirmed by other studies, the null effect of preceding disfluencies on relativizer omission would resemble Clark & Fox Tree's finding that uh and um contrast mainly in the length of the delays *following* them [6].

6. General discussion and conclusions

In earlier work with Thomas Wasow [17], we have presented preliminary evidence that relativizer presence is correlated with upcoming disfluencies. The studies presented here confirm that relativizers predict the presence of disfluencies in the NSRC *after controlling for other factors and speaker effects*. This finding is predicted by Ferreira & Dell's [7] availability-based hypothesis, according to which that is omitted when planning of the following material is finished. More generally, the results support the claim made in [7, 17, 21] that presence of that correlates with production difficulty.

Furthermore, the finding that relativizers are correlated with *more* disfluencies in the NSRC argues against the Alleviation Hypothesis. The available evidence suggests that relativizers do *not* help to alleviate production difficulties. On the contrary, the results support the Signal Hypothesis: relativizers are significant predictors of upcoming disfluencies even after properties of the modified NP as well as the disfluency rate before the NSRC are controlled for. In conclusion, while the current studies do not show that addressees *do* interpret optional relativizers as signals that the speaker is anticipating production difficulties, it shows that addressees *could* do so.

What are the consequences of these findings for the relation between audience design and optional word omission? The studies presented here show that optional that may be subject to audience design despite the fact that its distribution is governed by production pressures (in which case optional word omission would resemble uh/um, [6]). This is consistent with the observation that, although speakers do not seem to monitor their own speech for structural ambiguity, they use full-form clauses more frequently when addressees are present (compared to situations without addressees, [7]). This would be unexpected if optional that was uttered only to help speakers, but it is predicted by the Signal Hypothesis of optional word omission. If speakers use optional that to signal anticipated production difficulty, it makes sense that they use that more frequently in the presence of an addressee.

To end on a general note, the current finding argues that, for questions on the influence of audience design on morpho-syntactic variation (e.g., word omission, word order variation), it is misleading to focus exclusively on the avoidance of structural ambiguity. While current evidence suggests that speakers do not systematically avoid structural ambiguities (whether by prosodic phrase marking [1, 18], by insertion of disambiguating words [7, 24], or by choosing an unambiguous word order [1]), ambiguity avoidance is not the only way in which speakers can cater to addressees' needs: As shown by Clark & Fox Tree [6] for uh/um, speakers use collateral signals to keep addressees informed about the state of their production system.

7. Future work

More research on other word omission phenomena (e.g. complementizer omission, reduced subject-extracted relative clauses, omission of to after help) is necessary to see how general the finding presented here are.

With regard to the Signal Hypothesis, future research will decide whether addressees are sensitive to relativizer presence. Regardless of whether relativizers are signals or symptoms, a better understanding is needed of what kind of production difficulty that correlates with. For example, preliminary evidence suggests that relativizers also correlate with the presence of pauses immediately following them. Future research will show whether relativizers correlate specifically with suspensions of speech (as is the case for uh/um). Alternatively, speakers may use relativizers whenever they anticipate high workload. Ongoing work also investigates correlations of relativizers with the rate of speech before and within NSRCs.

8. Acknowledgements

I would like to express my heartfelt thanks to: Tom Wasow for continuous intellectual guidance; Herb Clark for discussions on the issue of audience design; Roger Levy and Joan Bresnan for statistical advice. Extraction of time information (that-length and pauses) would have been unfeasible without the incorporation of the Mississippi time-stamps into the Paraphrase Switchboard (by Shipra Dingare; LINK funding is gratefully acknowledged). I also owe thanks to Neal Snider, Dan Jurafsky, Laura Staum, Sasha Calhoun and Dave Orr for help with various aspects of this paper/project. None of the above necessarily agrees with the views presented here. This work has been funded in part by a Stanford Summer RAship.

9. References

- [1] Arnold, J. E., Wasow, T., Asudeh, A., & Alrenga, P. 2004. Avoiding Attachment Ambiguities: the role of Constituent Ordering. *Journal of Memory and Language*, 55(1), 55-70.
- [2] Bolinger, D. 1972. *That's that*. The Hague: Mouton.
- [3] Brennan, S. E., & Williams, M. 1995. The feeling of another's knowing: Prosody and filled pauses as cues to listeners about the metacognitive states of speakers. *Journal of Memory and Language*, 34, 383-393.
- [4] Bresnan, J., Carletta, J., Crouch, R., Nissim, M., Steedman, M., Wasow, T., et al. 2002. *Paraphrase analysis for improved generation*: HRCR Edinburgh-CLSI Stanford.
- [5] Clark, H. 2004. Pragmatics of language performance. In L. R. Horn & G. Ward (Eds.), *Handbook of pragmatics* (pp. 365-382). Oxford: Blackwell.
- [6] Clark, H. H., & Fox Tree, J. E. 2002. Using uh and um in spontaneous speaking. *Cognition*, 84, 73-111.
- [7] Ferreira, V. S., & Dell, G. S. 2000. Effect of Ambiguity and Lexical Availability on Syntactic and Lexical Production. *Cognitive Psychology*, 40, 296-340.
- [8] Fox, B. A., & Thompson, S. A. to appear. Relative Clauses in English conversation: Relativizers, Frequency and the notion of Construction. *Studies in Language*.
- [9] Fox Tree, J. E., & Clark, H. H. 1997. Pronouncing "the" as "thee" to signal problems in speaking. *Cognition*, 62, 151-167.
- [10] Goffman, E. 1981. Radio talk. In E. Goffman (Ed.), *Forms of talk* (pp. 197-327). Philadelphia, PA: University of Pennsylvania Press.
- [11] Hakes, D. T., Evans, J. S., & Brannon, L. L. 1976. Understanding sentences with relative clauses. *Memory & Cognition*, 4(3), 283-290.
- [12] Hawkins, J. A. 2001. Why are categories adjacent? *Journal of Linguistics*, 37, 1-34.
- [13] Hawkins, J. A. 2004. *Efficiency and Complexity in Grammars*. Oxford: Oxford University Press.
- [14] Jaeger, T. F., Fedorenko, E., & Gibson, E. 2005. Dissociation between Production and Comprehension Complexity. The 18th Annual CUNY Sentence Processing Conference, March 31st - April 2nd, 2005, Tuscon, AZ.
- [15] Jaeger, T. F., Orr, D., & Wasow, T. 2005. *Comparing Frequency-based and Complexity-based Accounts (of Relativizer Omission)*, The 18th Annual CUNY Sentence Processing Conference, March 31st - April 2nd, 2005, Tuscon, AZ.
- [16] Jaeger, T. F., & Wasow, T. 2005. *Processing as a Source of Accessibility Effects on Variation*. Paper presented at the Berkeley Linguistic Society.
- [17] Jaeger, T. F., & Wasow, T. 2005. *Production Complexity Driven Variation: The case of relativizer distribution in non-subject-extracted relative clauses*, The 18th Annual CUNY Sentence Processing Conference, March 31st - April 2nd, 2005, Tuscon, AZ.
- [18] Kraljic, T., & Brennan, S. E. 2005. Prosodic disambiguation of syntactic structure: For the speaker or for the hearer? *Cognitive Psychology*, 50, 194-231.
- [19] Levelt, W. J. M. 1983. Monitoring and self-repair in speech. *Cognition*, 14, 41-104.
- [20] Mateer, M., & Taylor, A. 1995. *Disfluency Annotation Stylebook for the Switchboard Corpus*. Unpublished manuscript, University of Pennsylvania.
- [21] Race, D. S., & MacDonald, M. C. 2003. *The use of "that" in the production and comprehension of object relative clauses*. Paper presented at the 26th Annual Meeting of the Cognitive Science Society.
- [22] Rayner, K., & Frazier, L. 1987. Parsing temporarily ambiguous complements. *Quarterly Journal of Experimental Psychology: Human Experimental Psychology*, 39A(4), 657-673.
- [23] Rohde, D. 2001. *Tgrep2 Manual*. Unpublished manuscript, Brain & Cognitive Science Department, MIT.
- [24] Roland, D., Elman, J. L., & Ferreira, V. S. in press. Why is that? Structural prediction and ambiguity resolution in a very large corpus of English sentences. *Cognition*.
- [25] Temperley, D. 2003. Ambiguity avoidance in English relative clauses. *Language*, 79(3), 464-484.
- [26] Venables, W. N., & Ripley, B. D. 2002. *Modern Applied Statistics with S* (Fourth edition ed.): Springer.
- [27] Wasow, T., & Jaeger, T. F. 2005. *Lexical Variation in Relativizer Frequency*, Expecting the unexpected: Exceptions in Grammar Workshop at the 27th Annual Meeting of the German Linguistic Association.