BLS 32, No 1 2006. DOI: http://dx.doi.org/10.3765/bls.v32i1.3453 (published by the Berkeley Linguistics Society and the Linguistic Society of America)

Phonological Optimization and Syntactic Variation: The Case of Optional *That*[*]

T. FLORIAN JAEGER
Stanford University

1. Introduction

The principle of Phonology-Free Syntax (see papers in Zwicky 1969) states that syntax should not be influenced by phonological constraints. This predicts that choice in syntactic variation should not be affected by phonological constraints, such as the avoidance of identical adjacent elements (Frisch et al. 2004; Leben 1973). Only a few studies have investigated the extent to which phonological encoding can influence syntactic encoding and they have come to conflicting results (see Bock and Eberhard 1993; Haskell and MacDonald 2003; MacDonald et al. 1993). I investigate the influence of phonological optimization at different levels of linguistic form on so called optional *that*, as in the following complement clauses (CC*s) and relative clauses (RC*s):[1]

(1a) You would think [$_{CC*}$ (*that*) there's no place for capital punishment in a civilized western country].
(1b) I heard [$_{CC*}$ (*that*) that's the second happiest day of your life].

(2a) I mean [$_{NP}$ any drugs [$_{RC*}$ (*that*) you want _]].
(2b) I mean [$_{NP}$ the way [$_{RC*}$ (*that*) it vibrates _]].

[*] I am grateful for the discussions that lead to this paper. I would like to thank D. Pesetsky, N. Richards, and A. Marantz for raising the question whether optional *that* is allomorphy rather than syntactic variation. I am particularly thankful to A. Antilla and L. Blumenfeld, as well as B. Labov, N. Snider, L. Staum, and D. Jurafsky for productive discussions and advice. None of the above necessarily agrees with the views expressed here. The work presented in this paper comes in parts from a larger study on optional *that* presented in detail in my thesis (Jaeger 2006a, 2010).

[1] Not all CCs and RCs exhibit optional *that* variation (see Huddleston and Pullum 2002). For most speakers of Standard American English, only finite, restrictive, non pied-piped, non-extraposed, non-subject-extracted RCs can occur without optional *that*. Similarly, some CC-embedding verbs have been claimed to require *that*. Here only CCs and RCs compatible with optional *that* are considered (henceforth CC* and RC*, respectively). The "[$_{RC*/CC*}$ …]" annotation merely highlights the constituent structure. No claims are intended about the constituent type of CCs and RCs.

Earlier studies have not distinguished between phonological constraints at different levels of linguistic form. For example, sentence phonology (prosody) is not the same as segmental phonology (see paper in Inkelas and Zec 1990), and there is evidence from speech errors that prosodic encoding precedes segmental encoding during language production (Cutler 1980). It is possible that the former would influence syntactic production but the latter not. Indeed, Zec and Inkelas (1990) present cross-linguistic evidence that this is the case. They propose that the only phonological information available to syntax is prosodic structure. I discuss three types of form-related OCP effects: (a) the potential avoidance of adjacent identical segments by producing or omitting optional *that* (Section 3); (b) the avoidance of adjacent identical word forms (Section 4); and (c) the avoidance of adjacent stressed/unstressed syllables (Section 5). The studies presented here also differ from earlier work in that they are based on a corpus of spontaneous speech rather than speech elicited in laboratory experiments, where participants (re-)produce sentences out of context (and often without a communicative goal). Before I turn to (a) – (c), I briefly provide an introduction to existing accounts of optional *that*.

2. Accounts of Optional *That*

The likelihood of optional *that* is co-determined by processing complexity (Ferreira and Dell 2000; Hawkins 2004; Race and MacDonald 2003).[2] For example, speakers are more likely to produce *that* for CC*s and RC*s that begin with complex subject phrases (Jaeger and Wasow 2006; Race and MacDonald 2003; Roland et al. 2005), for CC*s and RC*s that are less predictable given the preceding context (Jaeger 2006a; Wasow et al. 2011), and in disfluent environments (the two effects are cumulative, Jaeger 2006a).

In the linguistic literature, there is some controversy as to whether optional *that* is a case of allomorphy (Pesetsky 1991, inter alia) or syntactic variation (Doherty 2000; Ferreira and Dell 2000 inter alia). If optional *that* is a case of allomorphy, conditioning by phonological constraints would not be surprising. Pesetsky (1991) suggests that zero complementizers (the absence of optional *that* in CC*s) are affixes that attach to the embedding verb. The evidence given for this view is the apparent ban of zero complementizers in CC*s that are not adjacent to the embedding verb, as in (3) and (4) (from Richards 1997:297, 299):

(3) [$_{CC}$ *(That) the earth is flat] has been *proven* repeatedly.
(4) He didn't *say* to Imelda [$_{CC}$ *(that) he ate the mackerel].

This evidence is weaker than it may appear. Zero complementizers do not categorically *require* adjacency to the embedding verb. Examples like (4) *do*

[2] Some researchers (e.g. Dor 2005) have proposed that optional *that* correlates with a meaning difference. In Jaeger (2006:Ch. 1) I present arguments against at least the strongest version of such accounts (namely that *all* optional *that* variation is due to semantic differences). Ongoing work by R. Kinsey and T. Wasow investigates the extent to which optional *that* is driven by semantics.

occur without complementizer *that* (albeit rarely, see e.g. Hawkins 2001). Furthermore, both (3) and (4) are already accounted for by processing-based theories of optional *that*. Subject complement clauses, as in (3), are rare and hence low in predictability. Given that CC* predictability correlates with higher complementizer likelihood, predictability may provide an explanation both for apparently categorical cases like (3) and gradient effects in other environments. Similarly, processing-based accounts of optional *that* predict that CC*s with intervening material, as in (4), should be more likely to have a complementizer. So (3) and (4) are already captured by processing-based accounts that also capture gradient differences in optional *that* likelihood in other environments. Therefore the examples hardly present evidence for an allomorphy account of optional *that*.[3]

Note also that optional *that* does not exhibit unsystematic lexical gaps (another hallmark of allomorphy though not a necessary condition). While certain CC-embedding verbs have been claimed to require a complementizer, these verbs form systematic groups (defined by their semantics, e.g. manner of speaking verbs, and/or other properties, such as low subcategorization likelihood of a CC). Further potential evidence against an allomorphy account, comes from a phenomenon known as syntactic persistence, the tendency of speakers to re-use syntactic templates. Speakers are more likely to produce a CC*/RC* with *that* if they did so in preceding productions (Ferreira 2003; Jaeger 2006b). Crucially, the same holds for CC*s/RC*s without *that*. Compared to a baseline without any preceding CC*/RC*, speakers are more likely to omit optional *that* if they did so in preceding productions (ibid). Syntactic accounts readily predict the syntactic persistence effect on optional *that*.

I tentatively conclude that available evidence favors syntactic variation accounts of optional *that*. This makes optional *that* an interesting test case for the question to what extent syntactic production is affected by different types of phonological OCP constraints.

3. Segmental OCP

Applied to language production, Phonology-Free Syntax (Zwicky 1969) predicts that syntactic planning is not affected by phonological processing. So, under the assumption that optional *that* is a syntactic variation, no conditioning by phonological factors is expected. Next I describe how I use potential segmental OCP effects to investigate whether optional *that* is affected by segmental optimization.

A preference to avoid adjacent identical elements, the so-called Obligatory Contour Principle (henceforth OCP, Leben 1973, inter alia), has been documented for many levels of linguistic representation (for references and a brief overview, see Walter and Jaeger, 2008). The OCP does not only influence the organization

[3] Note that this is not to say that processing-based accounts account for the distribution of complementizers in *all* environments. For example, the fact that subject-extracted RCs seem to require a relativizer (in Standard American English) does not follow from processing-based accounts. Here it only matters that this fact does not follow from an allomorphy account either.

of grammar, it also affects choice in spontaneous speech. For example, *-t\d* deletion in spontaneous speech is much more *likely* before and after segments that share place and/or manner of articulation with *-t\d* (Strassel 2001), but it is not *required* in those environments. Frisch, Pierrehumbert, and Broe (2004:221) propose a unified account of gradient and categorical OCP effects, according to which the avoidance of adjacent similar elements eases processing.

As mentioned above optional *that* is correlated with processing complexity. Some researchers have argued that optional *that* serves to *ease* processing (Hawkins 2004; Race and MacDonald 2003). In that case, optional *that* may also be used to avoid adjacent identical segments. Consider the following example:

(5) Um, [$_{NP}$ the kinds of jobs [$_{RC*}$ *that* people need to be trained for now]] …

In (5), the final segment immediately preceding the optional *that* is /s/. It shares its manner of articulation (fricative) with the adjacent segment of the optional *that*. Similarly, the first segment immediately following the optional *that* shares its manner of articulation with the last segment of *that* (plosive). Theoretically, speakers could use optional *that* to avoid segmental OCP effects in two ways. Speakers could *omit* optional *that* whenever including it would lead to shared place and/or manner of articulation with at least one surrounding segment, as in (5). Speakers may also *insert* optional *that* if surrounding segments were to have identical place and/or manner of articulation, as in the following example:

(6) [$_{NP}$ The guy [$_{RC*}$ I saw yesterday]] …

I conducted separate corpus-based studies on the CC* and RC* data sets to test both of these hypotheses.

3.1 Data and Method

Given that I intend to test the effect of a phonological factor on optional *that*, a speech corpus was chosen as database. The Switchboard portion of the Penn Treebank corpora (release 3, Marcus et al. 1999) consists of 650 telephone dialogues on a variety of topics, yielding approximately 800,000 spoken words in over 100,000 turns. Using TGrep2 (Rohde 2005), 3,081 CC*s and 3,452 RC*s from the Switchboard were extracted.[4] All cases were annotated for a variety of factors known to influence optional *that* (for details, see Jaeger, 2006a:Ch. 3 & 4). After removing cases with incomplete variable information, the two databases

[4] CC*s introduced by *I guess*, *I think*, *I mean* or *I say* were excluded because these phrases often function as grammaticalized epistemic markers, in which case the "CCs" are not necessarily complement clauses (Cacoullos and Walker submitted; Thompson and Mulac 1991). This step removes over half of the data (analyses presented here return the qualitatively identical results after appropriate controls on the full data set). I also removed all cases for which the embedding verb did not occur often enough to determine whether the verb obligatorily requires a complementizer. Finally, RCs with *wh*-relativizers were excluded (for discussion, see Jaeger 2006a:Ch. 4.1).

contained 3,012 CC*s and 3,027 RC*s, respectively. About 32% of the CC*s and 41% of the RC*s in the database have optional *that*.

These cases were coded for three types of potential segmental OCP effects on optional *that*. For each case, it was determined (a) whether the immediately preceding segment (usually the embedding verb for CC*s and the head noun for RC*s) shared place and/or manner of articulation with the first segment of optional *that* (henceforth SIMILAR PRECEDING SEGMENT); (b) whether the immediately following segment (usually the first word of the CC*/RC* subject) shared place and/or manner of articulation with the last segment of optional *that* (henceforth SIMILAR FOLLOWING SEGMENT); and (c) whether the preceding and following segment had identical place and/or manner of articulation (henceforth SIMILAR SURROUNDING SEGMENTS). Since the Treebank Switchboard is not phonetically annotated, the Carnegie Mellon Pronouncing Dictionary (v0.6) was used to map words to their phonological form.[5] The 39 phonemes used by the Carnegie Mellon Pronouncing Dictionary (for a description, see *http://www.speech.cs.cmu.edu/cgi-bin/cmudict*) were divided into seven place categories ('alveolar', 'bilabial', 'dental', 'labio-dental', 'post-alveolar', 'velar', and 'vowel') and six manner categories ('approximant', 'lateral', 'fricative', 'nasal', 'plosive', and 'vowel'). Next I describe the distribution of the just mentioned phonological environment variables.

Table 1 – 2 (next page) summarize the distribution of SIMILAR PRECEDING SEGMENT in CC*s and RC*s, respectively. '*Same*' and '*different*' stands for 'same/different manner/place of articulation as the adjacent segment of *that*'. For CC*s, the preceding segment is always different in terms of its place of articulation, but there are 434 cases for which the preceding segment had the same manner of articulation as the first segment of *that*. For RC*s, identity in terms of the place of articulation is extremely rare, but the preceding segment frequently shares its manner of articulation with the first segment of *that* (869 cases).

Table 3 – 4 summarize the distribution of SIMILAR FOLLOWING SEGMENT in CC*s and RC*s, respectively. Absolute identity of preceding of following segments with potentially adjacent segments of optional *that* is extremely rare (see the '*same* - *same*' cells in Table 1 – 4). This means there are not enough data to test the effect of segmental *identity* on optional *that*. There are, however, enough data to test whether segmental *similarity* affects optional *that.* Frisch et al. (2004) present convincing evidence that OCP effects are gradient. Speakers do not only avoid adjacent identical elements, they also disfavor adjacent *similar* elements. I coded segments as similar if either the place or the manner of articulation or both were identical with the adjacent segment of *that*. This method yields 434 CC*s and 873 RC*s with potential OCP effects of the preceding segment, as well as 163 CC*s and 75 RC*s with potential OCP effects of the following segment.

[5] The caveat of this procedure is that the extracted phonological information only reflects what the phonological environment would look like if all words were fully articulated without being affected by surrounding phonological context, which they hardly ever are in spontaneous speech.

I also determined for each case whether the segment preceding the site of optional *that* shared place or manner of articulation with the segment following the site of optional *that*. Table 5 – 6 summarize the distribution of SIMILAR SURROUNDING SEGMENTS for CC*s and RC*s, respectively. Here '*same*' stands for identity in terms manner/place of articulation of the segment preceding *that* and the segment following *that*. As can be seen in Table 5, there are 232 CC*s for which the segments surrounding the site of optional *that* are identical with regard to *either* place *or* manner of articulation, as well as 359 CC*s where the surrounding segments are identical with regard to both place *and* manner of articulation. Similarly, there are 253 RC*s for which the surrounding segments are identical with regard one feature of articulation (see Table 6), and 235 RC*s for which the surrounding segments are identical with regard to both place and manner of articulation. Given these numbers, it is possible to code potential OCP effects of the surrounding segments as a three-way variable ('identity of both articulatory features' vs. 'identity of one articulatory feature' vs. 'no identity').

Table 1 Preceding segment (CC*s)

		PLACE	
		same	*diff.*
MANNER	*same*	0	434
	diff.	0	2,578

Table 2 Preceding segment (RC*s)

		PLACE	
		same	*diff.*
MANNER	*same*	0	869
	diff.	3	2,155

Table 3 Following segment (CC*s)

		PLACE	
		same	*diff.*
MANNER	*same*	29	56
	diff.	78	2,849

Table 4 Following segment (RC*s)

		PLACE	
		same	*diff.*
MANNER	*same*	8	42
	Diff.	25	2,952

Table 5 Surrounding segments (CC*s)

		PLACE	
		same	*diff.*
MANNER	*same*	359	201
	diff.	31	2,421

Table 6 Surrounding segments (RC*s)

		PLACE	
		same	*diff.*
MANNER	*Same*	235	242
	Diff.	11	2,539

If speakers omit optional *that* to avoid segmental OCP violations with the immediately preceding or following segment, optional *that* should be less likely if the segments was to share some articulatory feature with the adjacent segment of *that*. Also, if speakers insert optional *that* to avoid segmental OCP violations between the otherwise adjacent segments that immediately precede and follow the site of optional *that*, optional *that* should be more likely whenever those surrounding segments share at least one articulatory feature. Optional *that* should be most likely if those segments share both place and manner of articulation.

I present the results for CC*s and RC*s separately. In each case I start with simple omnibus tests. Then I present the results of regression analyses that control for other factors co-determining optional *that*.

3.2 CC* Results

Table 7 shows that optional *that* is much less likely (33.6%) if the preceding segment is similar to the first segment of optional *that* than if it is not (27.4% optional *that*). This correlation of SIMILAR PRECEDING SEGMENT and optional *that* likelihood is significant in the direction predicted by segmental optimization (Fisher's Exact $p < 0.02$). There also is a significant dependencies between optional *that* likelihood and SIMILAR FOLLOWING SEGMENT (see Table 8), but the effect is the opposite of the one expected by segmental optimization ($p < 0.0001$).

Table 7 Potential OCP with preceding segm.

		CC* *that*	
		N	%
Segmental Similarity	*no*	866	33.6%
	yes	119	27.4%

Table 8 Potential OCP with following segm.

		CC* *that*	
		N	%
Segmental	*no*	902	31.7%
	yes	83	50.9%

Table 9 Potential OCP of surrounding segments

		CC* *that*	
		N	%
Segmental Similarity	MANNER *and* PLACE	119	33.1%
	MANNER *or* PLACE	95	40.9%
	neither	771	31.8%

Finally, the effect of SIMILAR SURROUNDING SEGMENTS is hard to interpret. Overall, SIMILAR SURROUNDING SEGMENTS are associated with somewhat higher likelihood of optional *that* ($p < 0.05$), but the fact that similarity on only one dimension (MANNER or PLACE) is associated with a higher optional *that* likelihood than similarity on both dimension is unexpected. In sum, the initial Fisher Exact tests suggest a potential effect for SIMILAR PRECEDING SEGMENTS and maybe for SIMILAR SURROUNDING SEGMENTS.

To confirm the results with appropriate controls, SIMILAR PRECEDING SEGMENT, SIMILAR FOLLOWING SEGMENT, SIMILAR SURROUNDING SEGMENTS, and all control factors from Jaeger (2006a:Ch. 3 & 4) were entered into an ordinary logistic regression model of optional *that*.[6] The only segmental OCP effect

[6] To avoid problems due to clusters, the ordinary working model was bootstrapped 10,000 times with random replacement of speaker clusters. This procedure corrects for overly optimistic estimates (i.e. it guards against over-fitting). Stepwise backward factor elimination removed all factors from the bootstrapped model that do not contribute significant information (in terms of the change in data likelihood associated with their removal). For details I refer to Jaeger (2006a: Ch. 2). Another model was bootstrapped using replacement of clusters defined by the word immediately preceding the CC*. This correction prevents that effects associated with the embedding verb

that remained in the model is an inconsistent effect of SIMILAR SURROUNDING SEGMENTS. While similarity in terms of *either* MANNER *or* PLACE is associated with a 1.9 times increase in optional *that* likelihood (coefficient's $p < 0.0001$), similarity in terms of both features is associated with a 1.2 times decrease in the optional *that* likelihood ($p < 0.1$; against the predictions of a segmental optimization account). The effect associated with SIMILAR SURROUNDING SEGMENTS is therefore likely to be due to confounded. In sum, the CC* results do not provide evidence for the use of optional *that* for segmental optimization.

3.3 RC* Results

SIMILAR PRECEDING SEGMENTS and SIMILAR FOLLOWING SEGMENTS are associated with *higher* optional *that* likelihood (Fisher's Exact *p*s < 0.0001), contrary to the prediction of segmental optimization). The pattern for SIMILAR SURROUNDING SEGMENTS is again hard to interpret (see Table 12). While similarity of surrounding segments in terms of *either* MANNER *or* PLACE is associated with higher likelihood, similarity in terms of both MANNER and PLACE is associated with lower optional *that* likelihood than cases where the surrounding segments are dissimilar. If a simple measure of similarity (MANNER *or* PLACE identical?) is used, there is no dependency between SIMILAR SURROUNDING SEGMENTS and optional *that* likelihood ($p > 0.4$). In sum, prior to additional controls, the results do not suggest that optional *that* can be used for segmental optimization.

Logistic regression modeling (see footnote 6) revealed that most of the apparent effects are insignificant. The unexpected effect of SIMILAR PRECEDING SEGMENT, however, remains a significant factor in the opposite of the direction predicted by segmental optimization: A SIMILAR PRECEDING SEGMENT increases optional *that* likelihood by a factor of approximately 1.7 ($p < 0.0001$). In sum, again no evidence for the use of optional *that* for segmental optimization is found, but the significant effects in the unexpected direction will require further research before anything can be concluded with certainty. The available evidence, however, supports the Principle of Phonology-Free Syntax. Segmental optimization does not seem to affect (or at least not strongly affect) syntactic production.

---

are erroneously analyzed as segmental effects. Since both models are qualitatively identical, I do not discuss the latter any further (the same analyses, mutatis mutandis, were conducted for RC*s).

Table 10 Potential OCP with preceding segm.

		RC* *that*	
		N	%
Segmental Similarity	*no*	771	35.8%
	yes	480	55.0%

Table 11 Potential OCP with following segm.

		RC* *that*	
		N	%
Segmental	*no*	1,193	40.4%
	yes	58	77.3%

Table 11 Potential OCP of surrounding segments

		RC* *that*	
		N	%
Segmental Similarity	MANNER *and* PLACE	59	25.1%
	MANNER *or* PLACE	151	59.7%
	neither	1,041	41.0%

4. Word Form OCP

In Walter and Jaeger (2008), we presented evidence that optional *that* is affected by a word form OCP effect. Speakers avoid double *that* sequences at the beginning of CC*s and RC*s by omitting optional *that* when the following word is a pronoun or determiner *that* (henceforth deictic *that*), as in (1b) above:

(1b) I heard [$_{CC*}$ (*that*) that's the second happiest day of your life].

If an optional *that* was to directly precede a deictic *that*, speakers are more than twice as likely to omit the optional *that* than if it was followed by the word *the* or *this* (ibid). This word form OCP effect is unlikely to be due to prescriptive rules or editing. In Walter and Jaeger, we found that double *that* sequences are avoided in spontaneous speech as much as in formal writing. However, our studies did not include other control factors known to influence optional *that*. Therefore I included a word form OCP factor in the same regression models also used in the previous section.[7] There was only enough data to test the word form OCP effects for CC*s: 213 CC*s in my database (7%) begin with a deictic *that*. After controlling for all other factors, optional *that* is about 1.9 times less likely then expected if producing it would result in a double *that* sequence.

The word form OCP effect may be taken as evidence of phonological conditioning of optional *that*. Given the principle of Phonology-Free Syntax, this would argue against syntactic accounts of optional *that*. However, the word form OCP effect could also be related to lexical retrieval. The variant of a CC*/RC* with optional *that* may make retrieval of a following deictic *that* more difficult due to interference at the lemma level (i.e. when retrieving the lemma information for

[7] *All* results presented in this paper come from the same two models (one for the CC* data and one for the RC* data) that contained *all* OCP factors discussed in Section 3–5. For ease of presentation, I have chosen to present the results in separate sections.

deictic *that*). This slow-down associated with the processing of the variant with *that* may cause the variant without *that* to be chosen (see Ferreira and Dell 2000).

5. Rhythmic Optimization

Finally, I tested the effect of rhythmic optimization on optional *that*. For all CC*s and RC*s, the syllable immediately preceding and following the site of optional *that* was classified as either carrying primary stress or not.[8] Given that optional *that* is hardly ever stressed (Jurafsky et al. 2001) it is likely that it is usually included into one phonological phrase (e.g. a prosodic word) together with one of the words surrounding it. If speakers can use optional *that* to avoid stress lapses (adjacent unstressed syllables), optional *that* should therefore be dispreferred before and/or after unstressed syllables (depending on in which phrase the unstressed optional *that* would be included). Additionally, optional *that* maybe used to avoid a stress clash (adjacent stressed syllables).

The two stress factors (i.e. whether the preceding and following syllable are stressed) were entered into the regression model described in Section 3 (footnote 6). The results show that, both for CC*s and RC*s, only avoidance of lapses with *following* syllables seem to affect optional *that* likelihood. Optional *that* is about 2 times more likely before stressed syllables (coefficient's $p < 0.0001$ for CC*s; $p < 0.012$ for RC*s). No independent effect for the preceding syllable was found. This may indicate that optional *that* usually falls into the same phonological phrase as the first word following it. Preliminary support for this hypothesis comes from the fact that additional analyses revealed no evidence that speakers insert optional *that* even if this results in two adjacent stressed syllables. In other words, optional *that* seems to be entirely unaffected by the phonological properties of the preceding syllable. This would be expected if optional *that* is usually preceded by a prosodic break. Future research is necessary to determine whether this is the case.

6. Conclusions

Optional *that* is affected by avoidance of adjacent identical elements. While no evidence for *segmental* OCP effects on optional *that* were found, the word form OCP effect first reported in Walter and Jaeger (2008) was confirmed after appropriate controls. Since the word form OCP effects may be mediated via the lemma stratum, the evidence is compatible with a syntactic variation account of optional *that* (Doherty 2000; Ferreira and Dell 2000, inter alia) as well as with the

[8] Stress information was extracted from the Carnegie Mellon Pronouncing Dictionary (see Section 3.1). Personal pronouns, determiners, existential *there*, and prepositions were labeled as not carrying stress. There are two caveats to this procedure. First, the annotation does not reflect *actual* sentential stress but rather the *potential* of a syllable to receive sentential *stress*. Future work on prosodically annotated corpora is necessary to address this issue. Second, the distribution of stressed syllables is, unfortunately, somewhat correlated with the accessibility of the CC*/RC* subject (e.g. pronoun subject are unstressed and correlate with low relativizer likelihood). Although variation inflation factors did not indicate excessive colinearity (VIFs < 2.2), this confound has to be addressed in future work.

that (Doherty 2000; Ferreira and Dell 2000, inter alia) as well as with the principle of Phonology-Free Syntax (Zwicky 1969).

Finally, I have presented evidence that speakers also use optional *that* for rhythmic optimization (the avoidance of adjacent stressed or unstressed syllables). This extends earlier results from laboratory experiments that constituent ordering is affected by rhythmic optimization (MacDonald et al. 1993) to optional mentioning of function words in spontaneous speech. This result, too, is compatible with both syntactic variation accounts and the principle of Phonology-Free Syntax as long as the latter is restricted to word-level phonology. As mentioned earlier, there is evidence that prosodic phonology forms an independent level of representation (see paper in Inkelas and Zec 1990). Furthermore, evidence from speech errors suggests that prosodic encoding precedes segmental encoding: segmental errors are often due to confusion of the target word with a rhythmically similar word, but stress assignment errors do not seem to be driven by segmental similarity between words (Cutler 1980). Interestingly, Zec and Inkelas (1990) discuss cross-linguistic evidence that leads them to propose that prosodic structure is the *only* phonological information syntax has access to. The results presented here are consistent with this proposal: syntax is phonology-free except for prosody.

The fact that rhythmic optimization *co*-determines optional *that* along with many other factors (for an overview, see Jaeger 2006a, 2010; Roland et al. 2005) also underscores the fact that speakers' choices in sentence production are affects by constraints from a variety of sources.

References

Bock, J. K., and K. M. Eberhard. 1993. Meaning, Sound and syntax in English Number Agreement. *Language and Cognitive Processes* 8:57-99.

Cacoullos, Rena Torres, and James A. Walker. submitted. On the Persistence of Grammar in Discourse Formulas: A Variationist Study of that. *Linguistics*.

Cutler, Ann. 1980. Errors in Stress and Intonation. In Victoria Fromkin, ed., *Errors in Linguistic Performance: Slips of the Tongue, Ear, Pen, and Hand*, 67-80.

Doherty, Cathal. 2000. *Clauses Without 'That': The Case for Bare Sentential Complementation in English*. New York: Garland.

Dor, Daniel. 2005. Toward a semantic account of that-deletion in English. *Linguistics* 43:345–382.

Ferreira, V. S. 2003. The Persistence of Optional Complementizer Production: Why Saying "That" Is Not Saying "That" At All. *Journal of Memory and Language* 48:379-398.

Ferreira, Victor S., and Gary S. Dell. 2000. Effect of Ambiguity and Lexical Availability on Syntactic and Lexical Production. *Cognitive Psychology* 40:296-340.

Frisch, S. A., J. A. Pierrehumbert, and M. B. Broe. 2004. Similarity Avoidance and The OCP. *Natural Language and Linguistic Theory* 22:179-228.

Haskell, T.R., and M. C. MacDonald. 2003. Conflicting Cues and Competition in Subject-Verb Agreement. *Journal of Memory and Language* 48:760-778.

Hawkins, John A. 2001. Why Are Categories Adjacent? *Journal of Linguistics* 37:1-34.

Hawkins, John A. 2004. *Efficiency and Complexity in Grammars*. Oxford: Oxford University Press.

Huddleston, Rodney, and Geoffrey Pullum. 2002. *The Cambridge Grammar of the English Language*. Cambridge: Cambridge UP.

Inkelas, Sharon, and Draga Zec, eds. 1990. *The Phonology-Syntax Connection*. Chicago and London: The University of Chicago Press.

Jaeger, T. Florian, and Thomas Wasow. 2006. Processing as a Source of Accessibility Effects on Variation. In Cover, R.T and Kim Y. (eds.): *Proceedings of the 31st Annual Meeting of the Berkeley Linguistics Society*, 169-180. Ann Arbor, MN: Sheridan Books.

Jaeger, T. Florian. 2006a. Redundancy and Syntactic Reduction in Spontaneous Speech. Ph.D. thesis, Linguistics Department, Stanford University.

Jaeger, T. Florian. 2006b. Syntactic Persistence in Real Life (Spontaneous Speech). Paper presented at *The 19th Annual CUNY Sentence Processing Conference*.

Jaeger, T. Florian. 2010. Redundancy and Reduction: Speakers Manage Syntactic Information Density. *Cognitive Psychology*, 61(1), 23-62.

Jurafsky, Dan, A. Bell, M. Gregory, and W. D. Raymond. 2001. Probabilistic Relations Between Words: Evidence from Reduction in Lexical Production. In Joan Bybee and Paul Hopper, eds., *Frequency and the Emergence of Linguistic Structure*. Amsterdam: Benjamins.

Kinsey, R., Wasow, T., and Jaeger, T. F. 2007. What does "that" mean? Experimental Evidence against the Principle of No Synonymy. 81st Annual Meeting of the Linguistic Society of America, Anaheim.

Leben, Will. 1973. Suprasegmental Phonology. Ph.D. diss., MIT.

MacDonald, M.C., J. K. Bock, and M. H. Kelly. 1993. Word and World Order: Semantic, Phonological, and Metrical Determinants of Serial Position. *Cognitive Psychology* 25:188-230.

Marcus, Mitchell P., Beatrice Santorini, Mary Ann Marcinkiewicz, and A. Taylor. 1999. *Treebank III*. University of Pennsylvania: Linguistic Data Consortium.

Pesetsky, David. 1991. Zero Syntax II. Ms. MIT.

Race, David S., and Maryellen C. MacDonald. 2003. The Use of "That" in the Production and Comprehension of Object Relative Clauses. Paper presented at *26th Annual Meeting of the Cognitive Science Society*.

Rohde, D. 2005. Tgrep2 Manual. Ms., MIT.

Roland, Douglas, Jeffrey L. Elman, and Victor S. Ferreira. 2005. Why is That? Structural Prediction and Ambiguity Resolution in a Very Large Corpus of English Sentences. *Cognition*:1-28.

Strassel, Stephanie. 2001. Corpus-Based Sociolinguistic Research: A -t/d Deletion Case Study. Paper presented at *NWAV 30*.

Thompson, Sandra A., and Anthony Mulac. 1991. The Discourse Conditions for the Use of Complementizer 'That' in Conversational English. *Journal of Pragmatics* 15:237- 251.

Walter, Mary Ann, and T. Florian Jaeger. 2008. Constraints on English that-drop: A strong lexical OCP effect. In Edwards, R.L., Midtlyng, P.J., Stensrud, K.G., and Sprague, C.L. (eds.) *Proceedings of the Main Session of the 41st Meeting of the Chicago Linguistic Society*, 505-519. Chicago, IL: CLS.

Wasow, Thomas, T. Florian Jaeger, and David Orr. 2011. Lexical Variation in Relativizer Frequency. In H. Simon and H. Wiese (eds.): *Proceedings of the 2005 DGfS workshop "Expecting the unexpected: Exceptions in Grammar"*, 175-196. Berlin/New York: De Gruyter Mouton.

Zec, Draga, and Sharon Inkelas. 1990. Prosodically Constrained Syntax. In Sharon Inkelas and Draga Zec*,* eds., *The Phonology-Syntax Connection*. Chicago and London: Chicago University Press.

Zwicky, Arnold M. 1969. Phonological Constraints in Syntactic Descriptions. *Papers in Linguistics* 1:411-463.

T. Florian Jaeger
University of Rochester
Human Language Processing Lab
Brain and Cognitive Sciences
Computer Science
Box 270268
Rochester, NY 14627-0268

fjaeger@bcs.rochester.edu