Processing as a Source of Accessibility Effects on Variation

T. FLORIAN JAEGER & TOM WASOW
Stanford University

0 Introduction*

English restrictive non-subject-extracted relative clauses (i.e. relative clauses in which the extracted element is *not* the subject of the relative clause; henceforth NSRCs) exhibit variation in that the relativizer can be omitted:[1]

(1) This is the first president$_i$ (that) nobody voted for _$_i$.

A variety of factors are known to influence relativizer likelihood (see, inter alia Biber et al. 1999; Fox and Thompson in review; Tagliamonte, Smith, and Lawrence 2005; Temperley 2003; Tottie 1995). We present new evidence that the *conceptual accessibility* (Bock and Warren 1985:50) of an NSRC's subject affects relativizer likelihood: The more accessible the referent of a NSRC's subject is in working memory, the less likely the NSRC is to have a relativizer. We link this finding to research on the production and comprehension of relative clauses, and so integrate the observed accessibility effect into a uniform processing account of relativizer variation (Race and MacDonald 2003; Jaeger and Wasow 2005).

In Section 1, we show that relativizer omission is sensitive to the *derived accessibility* (Prat-Sala and Branigan 2000) of the NSRC's subject – that is, the subject referent's salience/givenness in discourse. In Section 2, we outline a processing-based account of the observed effects. In Section 3, we show that relativizer variation is also affected by the *inherent* accessibility of the NSRC's subject, specifically number and referentiality. Section 4 concludes with the consequences for future research and a brief summary of the observed effects.

* We would like to thank G. Bouma, S. Calhoun, E. Gibson, R. Katzir, D. Orr, S. Vasishth, the BLS 31 audience, and especially D. Jurafsky and N. Snider for useful feedback. Many thanks to D. Rohde for help with technical questions. We are grateful to the Edinburgh-Stanford LINK Paraphrase project for access to the Paraphrase Switchboard, and to TedLab, MIT for providing one of the authors (TFJ) with a stimulating work environment during the early stages of this project.

[1] Several types of NSRCs do not exhibit relativizer variation and were therefore excluded from the study presented below.

1 Derived Accessibility and Relativizer Variation

Derived accessibility is due to a referent's salience/givenness in discourse. It depends on the context of use and is hence not inherent to the referent. A referent's derived accessibility has long been noted to affect the choice of linguistic expression referring to it (Ariel 1990; Arnold 1998; Givón 1983; Gundel, Hedberg, and Zacharski 1993; for a recent overview, see Ariel 2001). In the accessibility scale in (2), based on Ariel (1990:73), expressions higher on the scale refer to referents that are more salient or more recently mentioned in the discourse.

(2) Pronoun > Demonstrative > First Name > Definite NP > Indefinite NP

Although past research on relativizer variation has not invoked accessibility, several works have noted correlations between relativizer likelihood and properties of the NSRC's subject NP (Biber et al. 1999; Fox and Thompson in review; Temperley 2003; Tottie 1995). The summary of earlier studies in Table 1 suggests that the more accessible an NSRC's subject expression is, the less likely is a relativizer.

	Pronoun		Lexical NP		
Subject expression:	**1st.SG**	**Other**	**Proper N**	**Def. NP**	**Indef. NP**
Tottie (1995)	23%			48%	70%
Temperley (2003:475)	11%		37%	45%	
Biber et al. (1999:620)	30-40%		80-95%		
Fox & Thompson (in review)	34%	41%	55%		

Table 1 – Relativizer Frequency by Subject NP Type of NSRC

Although earlier studies suggest that the NSRC's subject affects relativizer likelihood, they were all conducted on rather small data sets.[2] This data sparseness limited earlier studies to two- or three-way distinctions with regard to the NSRC's subject. Our investigations confirm and extend earlier results to a finer-grained distinction of accessibility. Another limitation of most earlier studies is that they relied entirely on *written* data. Studying relativizer variation in *spoken* language has two advantages. First, informal spoken language is less subject to prescriptive influences, a potentially confounding factor Second, we ultimately are interested in whether the observed effects are due to processing (see Section 2) and any processing effects will show up more clearly in naturally occurring spontaneous speech, since spoken language is subject to real time processing pressures.[3] Fox

[2] Tottie's study is based on 385 NSRCs from the British National Corpus. Temperley extracted 329 NSRCs from the Wall Street Journal. Fox and Thompson analyzed 195 NSRCs from 36 audio- and video-taped conversations of friends and family.

[3] One may object that a well-controlled production experiment may be more suited for the question at hand. However, the corpus-based approach taken here has the advantage of ecological validity. Thus we believe that corpus-based studies complement experimental studies.

and Thompson (in review) is the only study we have found on relativizer variation in speech that investigated effects of the NSRC's subject type, and it is based on fewer than 200 NSRCs. To construct a sufficiently large database, we automatically extracted all 4,405 NSRCs from the Paraphrase Switchboard corpus of informal spoken English.[4] Of these, 698 (25%) are introduced by a *wh*-relativizer. Of the remaining 3,707, 56.6% contain a *that* relativizer and 43.4% have no relativizer. Inspection revealed that many of *wh*-relativizers in our sample are not optional (e.g., because the relative clause is non-restrictive), so the quantitative studies reported on here are based on comparing NSRCs lacking relativizers with those introduced by *that*.[5]

About 22% of the NPs in the Paraphrase Switchboard are annotated for givenness. As Table 2 shows, given subjects correlated with a significantly lower relativizer frequency than non-given subjects (χ^2= 16.6, p< 0.001, N= 1,042).[6]

		Total	Relativizer Frequency
Givenness (of subject expression)	Given	884	51.5%
	Not given	158	69.9%
Total		1,042	χ^2= 16.6, p< 0.001

Table 2 – Givenness of an NSRC's Subject and Relativizer Frequency

This result is encouraging but our interest is in whether relativizer variation is sensitive to *degrees* of accessibility (see the discussion in Ariel 2001:37f.), as predicted by a processing-based account of accessibility (see Section 2). To address this question, we used the type of an NSRC's subject expression as an indicator of the subject's derived accessibility.

We grouped the NSRCs in our database into six classes based on the NSRC's subject expression: 1st (*I*, *we*), 2nd (*you*), and 3rd person pronouns (*he*, *she*, *it*, *they*), NPs introduced by a possessive pronoun (e.g. *my kids*), definite NPs (introduced by *the*, e.g. *the woman*), and indefinite NPs (introduced by *a(n)*, e.g. *a teacher*). Table 3 summarizes the average relativizer frequency for each of the subject types. Over 92% of all NSRC subjects in our database fall into one of these six groups. The remaining NSRC subjects were either a demonstrative pronoun (e.g. *that*, *these*; 15 cases), a proper name (30), a one-word quantifier NP (e.g. *someone*; 26 cases), the pivot of an existential construction (22), or a lexical NP which was not introduced by *the*, *a(n)*, or a possessive pronoun. The last

[4] We used the Paraphrase Switchboard (Bresnan et al. 2002), which contains the same conversations as the Treebank III Switchboard release (Marcus et al. 1999). The Paraphrase corpus is annotated for animacy (Zaenen et al. 2004), and in parts for information structure, referentiality, and co-reference (Nissim et al. 2004). The corpus consists of 650 transcribed telephone conversations between two strangers (on a list of selected topics) totalling approximately 800,000 words.

[5] We also ran all of our tests using the larger database that included the *wh*-relativizer examples. For the effects reported here, the results were all qualitatively the same.

[6] Here, 'given' refers to referents that have been explicitly mentioned in the preceding dialogue.

group contains NPs introduced by demonstrative determiner (e.g. *those ten people*; 12 cases), some kind of quantifier (e.g. *most of the countries*; 23 cases), as well as bare NPs (78 generics like *politicians* and 10 mass nouns like *copy room paper*). Since for all these groups there were either no clear predictions in terms of derived accessibility (e.g. bare NPs) or the groups were too small and/or too heterogeneous with regard to accessibility (e.g. quantified NPs), we do not discuss them here.[7]

Subject expression	**Total**	**Relativizer Frequency**
1st person pronoun	1,905	39.7%
2nd person pronoun	571	42.9%
3rd person pronoun	762	43.4%
Possessive NP (with possessive pronoun)	70	47.8%
Definite NP	97	54.6%
Indefinite NP	18	77.8%
Total	3,423	$\chi^2(5)= 21.8, p< 0.001$

Table 3 – NSRC's Subject and Relativizer Frequency in the Switchboard

As can be seen in Table 3, relativizer likelihood increases the less accessible the subject of the NSRC is. Admittedly, the numbers get rather small towards the bottom of the table, e.g. for indefinite NPs. Taken together with results by others (see Table 1 above) our results nevertheless provide strong support for the hypothesis that derived accessibility influences relativizer likelihood. Furthermore, our results confirm earlier findings (e.g. Prat-Sala and Branigan 2000:180) arguing that accessibility is not a binary but a gradient property.

Could the observed effects be due to the morphosyntactic complexity of the subject (e.g., its length in phonemes, syllables, words, or its weight in syntactic nodes)? In order to address this question, we measured the length (in words) of all NSRC subjects.[8] We found the following. First, even one-word lexical subject NPs have a significantly higher relativizer rate (65.1%) than pronominal subject NPs (41.6%; $\chi^2= 19.1$, $p< 0.001$, N= 3,361). Second, NSRCs with lexical one-word subjects actually have a slightly though not significantly *higher* relativizer frequency (57.3%) than NSRCs with multi-word subjects ($\chi^2< 1.6$, $p> 0.2$, N= 339). We conclude that the accessibility effect observed above is independent of the grammatical weight of an NSRC's subject. We turn next to the question of *why* this should be the case.

[7] For the effect of *that XP*/*that*-subjects on relativizer/complementizer omission, see Walter & Jaeger (2005). E.g., NSRCs with *that* pronoun subjects exhibit extremely low relativizer frequency (18.2%). Walter & Jaeger attribute this to the lexical Obligatory Contour Principle.

[8] For an overview of measures of grammatical weight, see the discussion in Wasow (2002:23-32). Wasow provides evidence that existing measures of grammatical weight are so highly correlated that it is virtually impossible to tease them apart (cf. Szmrecsányi 2004). It is therefore likely that our result extends to other purely syntax-based measure of an NSRC's subject complexity.

2 Accessibility affects are explained by processing complexity

Following others (Hawkins 2001, 2004; Race and MacDonald 2003; see also Jaeger and Wasow 2005), we propose a processing-based account of the correlation between accessibility and relativizer variation:

[A] The more accessible a subject referent is in working memory, the easier it is to construct mental representations of it.
[B] Constructing a mental representation of the subject referent accounts for a significant amount of the overall processing complexity of an NSRC.
[C] The lower the processing complexity associated with an NSRC, the less likely is a relativizer.[9]

Such an account has multiple advantages. First, accessibility effects on relativizer variation are attributed to a uniform source: processing complexity. This avoids the problematic claim (Fox and Thompson in review; Jespersen 1922) that there is a meaning difference linked to relativizer variation. Race & MacDonald (2003) provide convincing evidence that the absence or presence of a relativizer does *not* correlate with a meaning difference (based on similar work on complementizer omission by Ferreira & Dell 2000). Moreover, there is independent evidence for [A] to [C].

Independent evidence for [A] comes from the sentence production literature. While details about *how* accessibility affects word order in production are still unresolved, a rich body of research on many languages shows that highly accessible subjects are produced more rapidly than subjects that are low on the accessibility scale (for a recent literature overview, see van Nice and Dietrich 2003).[10] [A] is also supported by the fact that high frequency forms exhibit faster lexical retrieval (accessibility correlates with occurrence frequency, cf. Table 1).

Independent evidence for [B] comes from the *comprehension* of NSRCs. Warren & Gibson (2002) present a series of self-paced reading experiments showing that decreased accessibility of an NSRC's subject increases reading times on the following verb. For example, one of their experiments (p. 86ff.) shows a monotonic increase in reading times on the embedded verb, e.g. *praised* in (3), as the accessibility of the subject of a double nested NSRC decreases (from *you* to *a reporter*). Thus the accessibility of an NSRC's subject influences processing.

(3) The old lady who the government assistance program which *you*/*Bill*/*the reporter*/*a reporter* praised had saved did not have enough money …

[9] Here we are not concerned with whether speakers insert relativizers to facilitate production or to assist listeners' comprehension. We have, however, argued elsewhere that relativizer variation is primarily driven by *production*-complexity (Jaeger and Wasow 2005; see also Race and MacDonald 2003; see Ferreira and Dell 2000 on a similar analysis for complementizer omission).

[10] Ferreira (Ferreira 1994) presents evidence that accessibility effects on word order are (partly) mediated via thematic role assignment (see van Nice and Dietrich 2003 for an overview). Thus [A] should be taken to primarily apply to proto-typical subject roles (i.e. agentive subjects).

Finally, independent evidence for [C] comes from several studies arguing that relativizer likelihood is influenced by a variety of processing-related factors such as the amount of intervening material between the head noun and the beginning of the relative clause (Jaeger, Orr, and Wasow 2005; Quirk 1957), the overall complexity of the NSRC (Race and MacDonald 2003; Jaeger, Orr, and Wasow 2005), the predictability of the NSRC (Wasow and Jaeger in progress; Jaeger, Orr, and Wasow 2005), and ambiguity avoidance (Temperley 2003). Example (4) illustrates the effect of intervening material; (5) demonstrates the effect of the NSRC's predictability (the more likely the NSRC due to e.g. uniqueness requirements of the definite article, and/or a superlative, the less likely is a relativizer).

(4) … [the other problem with capital punishment [$_{\text{NSRC}}$ (that) you run into]] …
(5) a. Tell me about [a movie [$_{\text{NSRC}}$ (that) you saw]].
b. Tell me about [the movie [$_{\text{NSRC}}$ (that) you saw]] …
c. Tell me about [the last movie [$_{\text{NSRC}}$ (that) you saw]] …

Similar results have been obtained for complementizer omission (e.g. Ferreira and Dell 2000), where complementizer likelihood correlates positively with the overall complexity and likelihood of the complement clause.

The independent motivations for [A] to [C] make the analysis of the accessibility effects outlined above a desirable explanation for relativizer omission. In a nutshell, we claim that the effects observed in Section 1 are due to faster construal of accessible NSRC subject referents in working memory. If this is correct, then other properties known to influence construal of referents should also affect relativizer variation. The *inherent* accessibility of referents (Prat-Sala and Branigan 2000) is such a property.

3 Inherent Accessibility and Relativizer Variation

The factors contributing to inherent accessibility are features that make a referent easier to construct for participants in a conversation independent of the context of the conversation (e.g. because the reference is conceptually less complex or because the type of reference is more frequently employed). There may well be many such factors, but we will discuss only three here: number, referentiality, and animacy.

3.1 Number

Referents of singular NPs are inherently more accessible than referents of plural NPs (we assume that, ceteris paribus, the construction of multiple referents in working memory is more complex than the construction of a single referent). And, as predicted, plural referents correlate with significantly higher relativizer likelihood (49.1%) than singular referents (38.3%; $\chi^2 > 30$, $p < 0.001$). This effect also holds separately within pronouns ($\chi^2 = 23.4$, $p < 0.001$, $N = 2{,}671$) and common nouns ($\chi^2 = 5.7$, $p = 0.02$, $N = 298$).

3.2 Referential vs. Impersonal Uses of Pronouns

The pronouns *you*, *we*, and *they* can be used either to refer to specific individuals or impersonally, as in (6).

(6) a. But, uh, they have sort of like, uh, things [$_{NSRC}$ that *you*'re not like reimbursed for …]
b. And one way [$_{NSRC}$ that *we* do it sort of in Iowa is that we can take some of our clothes to the consignment shops].
c. I don't remember what they call it … some kind of word [$_{NSRC}$ *they* use when you get a positive indication of drugs] …

Impersonal references are inherently less accessible (see Ariel 2001:68 and references therein) and should therefore incur a higher processing load.[11] As predicted by a processing account impersonal uses of pronouns correlate with significantly higher relativizer likelihood (52.4%) than referential uses (39.9%; χ^2= 9.5, p< 0.01, N= 465) in the portion of the Paraphrase Switchboard annotated for referentiality (about a fifth of the corpus; see Nissim et al. 2004).

3.3 Animacy

Approximately 94% of the NSRCs in the database were annotated for animacy (see footnote 4). We investigated the effect on relativizer likelihood of human vs. inanimate NSRC subject referents (excluding a third category containing animals and organizations due to the small number of observations). Since the referential status (e.g. pronoun vs. common noun) of an NSRC's subject affects relativizer likelihood, as do its person and number, we are left with four possible test domains for animacy effects (in order to avoid confounds): singular and plural common nouns, and singular and plural 3rd person pronouns.

However, in our data, referents of 3rd person plural pronoun subjects (*they*) are overwhelmingly animate (96.6%); hence we did not have enough data to test for an animacy effect in that category (only 11 uses of *they* referred to inanimates). Unfortunately, comparisons for 3rd person singular pronouns (i.e. *he*, *she*, *it*) are likewise problematic since a great many uses of *it* (43%) occur in the idiomatic string … *the way it* …:

(7) … the way [$_{NSRC}$ it is/was/goes/has to be …

Such collocations almost categorically occur without a relativizer: 96.6% of all combinations of *way* as a head noun and *it* as the NSRC subject do not have a relativizer. Since all instances of examples like (7) are *annotated* as inanimates in the Paraphrase Switchboard, although for most of these cases, it is questionable

[11] It may be that impersonal references are less accessible (in part) because they are usually not anaphoric (and therefore not given). In that case, referentiality effects would (in part) be due to *derived* accessibility.

whether *it* refers to anything at all, this creates a strong confound *against* an animacy effect. Further complicating matters, NSRCs with *it* subjects modify semantically light nouns such as *time, place, thing, way,* etc. in 59.1% of all cases, whereas NSRCs with *he* or *she* subjects modify such light nouns in only 34.6% of all cases. We have shown elsewhere that light head nouns strongly favor NSRCs without a relativizer (Wasow and Jaeger in progress). Hence, unless NSRC subject animacy is an extremely strong predictor of relativizer absence, we would expect *fewer* relativizers for NSRCs with *it* subjects. This is indeed the case (χ^2= 30.2, p< 0.001). To conclude, looking at NSRCs with 3rd person singular pronoun subjects, there are not enough NSRCs with semantically heavy head nouns (not favoring relativizer omission) for a potential animacy effect to surface.

This left us with lexical NSRC subjects. Since lexical plural subject are less likely to be inanimates (16.4%) than lexical singular subjects (38.8%) and NSRCs with singular subjects are less likely to occur without a relativizer (see above), we examined animacy for singular and plural referents separately. Surprisingly, both groups seem to exhibit an *anti*-animacy effect: inanimate subject referents correlate with *lower* relativizer likelihood (40.3% for singular and 29.0% for plural referents) than human referents (70.4% for singular and 65.2% for plural referents; χ^2s > 10, Ps< 0.005, N= 133 for singular and N= 125 for plural referents).

However, this effect is severely confounded. There is a strong correlation between the grammatical function of the extracted element in the NSRC and the animacy of the NSRC's subject. And the grammatical function of the extracted element is strongly correlated with relativizer likelihood. That is: NSRCs in which the extracted element is an adverb, as in (8), are much less likely (25.2%) to have a relativizer than NSRCs in which the extracted element is an object (76.5%). This is in turn due to the high ratio of the above-mentioned semantically light head nouns (e.g. *time, place, way*) for NSRCs with adverb gaps.

(8) a. … every time [$_{NSRC}$ her sons starts taking things for granted] …
b. … the way [$_{NSRC}$ our state tax is here] …

Crucially, NSRCs with adverb gaps are also far more likely to have an inanimate lexical subject (78.2% of all cases) than NSRCs with an object gap (only 19.9% of which have an inanimate lexical subject). Once this confound is controlled for, no animacy or anti-animacy effect remains (all χ^2s < 1.2).

3.4 Discussion

A rich literature on speakers' choice in production shows that speakers prefer to utter highly accessible referents early in the sentence (for a recent overview, see van Nice and Dietrich 2003). While to the best of our knowledge most of this literature has focused on matrix clauses, there is some evidence that similar effects show up during the production of embedded clauses. For example, Gennari et al. (2005) show that, in object-extracted relative clauses with inanimate head nouns, speakers prefer to produce animate agents early (i.e. as subjects).

Unfortunately, our database does not include enough examples of inanimate NSRC subjects to allow a meaningful investigation of the effect of animacy on relativizer realization. We leave this issue open for future research. The prediction of a processing-based analysis of relativizer variation is clear. If carefully controlled for other factors, a large enough data set of NSRCs should exhibit an animacy effect on relativizer likelihood.

Even though we did not find an animacy effect, support for a processing-based account comes from two rather clear effects of inherent accessibility: Both the number and the referentiality of the NSRC's subject pronoun have the predicted effect on relativizer likelihood: the more accessible the NSRC's subject referent is, the less likely is a relativizer. More support for the hypothesis that the inherent complexity of referents influences relativizer likelihood comes from additional comparisons we conducted. Although too small for meaningful statistical analysis (see Section 1), generic, mass noun, and quantified lexical NP subjects in our dataset (all arguably conceptually complex) correlate with high relativizer frequencies (63.2% to 66.7%), as we predict. Similarly, quantifier subjects like *everybody*, *anybody*, or *someone else* correlate with high relativizer frequency (72.3%) even though most of them are one-word expressions.

Finally, note that NSRCs with an expletive *it* subject, as in (7), or an existential *there*, as in (9), almost never have a relativizer (e.g. only 27.3% of the 22 NSRC with an existential *there* have a relativizer).

(9) … anytime [$_{\text{NSRC}}$ there is a change in weather …]

Under the assumption that relativizer presence is primarily correlated with processing difficulties at the beginning of an NSRC: the low relativizer frequency correlated with non-referring NSRC subject expressions, if confirmed on larger datasets, would provide further evidence for hypothesis [A] in Section 2: If the first word of an NSRC does not refer to anything (i.e. is conceptually less complex than a referring expression), it is easier to construct/retrieve that expression from working memory.

4 General Discussion and Conclusions

Although much research on word order variation (e.g., Wasow 2002; Hawkins 2004) has noted that some factors associated with accessibility (such as grammatical weight, definiteness, pronominality, and givenness) influence word order, almost nobody has explicitly linked these findings to conceptual accessibility (Bock and Warren 1985). One of the few exceptions, Bresnan et al. (2005) show that speakers are more likely to choose the double object variant of the dative alternation, when the recipient is more accessible.

The findings presented here argue that accessibility affects not only word *order* but also word *omission* variation. Both cases of variation have in common that highly accessible forms occur earlier in the sentence (here due to the omission of the relativizer). The sentence production literature (e.g. Bock and Warren 1985)

already offers a plausible explanation for this fact. Ceteris paribus, formulation of highly accessible referents takes less time. Thus, for highly accessible subjects, omitting a relativizer actually can save time (and therefore be efficient), whereas omission would not buy any time if the formulation of the subject expression has not been finished. More generally, we incorporated the observed accessibility effects into a processing account (Section 2) and showed that the predictions of such an account are at least partly supported by effects of the inherent accessibility of an NSRC's subject.

We have also shown that the accessibility effect cannot be reduced to grammatical weight. As a matter of fact, an NSRC subject's grammatical weight does not seem to contribute to relativizer variation after its accessibility is controlled for. Type frequency (which is highly correlated with accessibility, cf. Table 1), on the other hand, we for now subsume as a factor contributing to accessibility.

Using a large database enabled examination of more subtle accessibility effects on relativizer variation, which support the hypothesis that derived accessibility is a gradient phenomenon (see Ariel 2001:37f. for references making claims for or against this hypothesis). The accessibility-based account proposed in Section 2 offers a uniform analysis of the variation in relativizer likelihood associated with different subject expressions (observed here and in earlier research; cf. Table 1 in Section 1) *and* the variation associated with the givenness of the subject (and is as such to be preferred over accounts that treat givenness as a binary factor, e.g. Temperley 2003).

Finally, a processing-based account of the accessibility effects raises an intriguing possibility. Integrative approaches to variation (e.g., Hawkins 2004; Wasow 2002) investigate the extent to which variation is due to e.g. processing. The underlying idea is that, whenever speakers have a choice (as defined by the grammar), they structure utterances so as to minimize processing complexity. According to Hawkins (2004), such preferences then eventually lead to cross-linguistic variation. In the current case, the suggested link between accessibility and processing complexity connects to cross-linguistic variation in case-marking, e.g. phenomena like Differential Case Marking (DCM, e.g. Aissen 2003). Just as case-marking in languages with DCM signal subject referents low in accessibility and/or object referents high in accessibility, relativizers may signal NSRC subjects low in accessibility and, more generally, NSRCs that are hard to process.

5 References

Aissen, Judith. 2003. Differential Object Marking: Iconicity vs. Economy. *Natural Language and Linguistic Theory* 21:435-83.

Ariel, M. 1990. *Accessing Noun Phrase antecedents*. London: Routledge.

———. 2001. *Accessibility Theory: An Overview*. Edited by T. Sanders and al., *Text representation: Linguistic and psycholinguistic aspects*. Amsterdam: Benjamins.

Arnold, Jennifer E. 1998. Reference form and discourse patterns, Department of Linguistics, Stanford University.

Biber, Douglas, Stig Johansson, Geoffrey Leech, Edward Finegan, and Susan Conrad. 1999. *Longman grammar of spoken and written English*. London: Longman.

Bock, J. K., and R. K. Warren. 1985. Conceptual Accessibility and Syntactic Structure in Sentence Formulation. *Cognition* 21 (1):47-67.

Bresnan, Joan, Jean Carletta, Richard Crouch, Malvina Nissim, Mark Steedman, Tom Wasow, and Annie Zaenen. 2002. Paraphrase analysis for improved generation. In *LINK project*: HRCR Edinburgh-CLSI Stanford.

Bresnan, Joan, Anna Cueni, Tatiana Nikitina, and Harald Baayen. 2005. Predicting the Dative Alternation. Paper read at Royal Netherlands Academy of Science Workshop on Cognitive Foundations of Interpretation.

Ferreira, Fernanda. 1994. Choice of passive voice is affected by verb type and animacy. *Journal of Memory and Language* 33 (6):715-36.

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

Fox, Barbara A., and Sandra A. Thompson. in review. Relative Clauses in English conversation: Relativizers, Frequency and the notion of Construction. *Studies in Language*.

Gennari, Silvia, Jelena Mirkovic, and Maryellen C. MacDonald. 2005. The role of animacy in relative clause production. The 18th Annual CUNY Sentence Processing Conference, March 31st - April 2nd, 2005, Tuscon, AZ.

Givón, Talmy. 1983. *Topic continuity in discourse : a quantitative cross-language study*: John Benjamins.

Gundel, Jeanette K., Nancy Hedberg, and Ron Zacharski. 1993. Cognitive status and the form of referring expressions in discourse. *Language* 69 (2):274–307.

Hawkins, John A. 2001. Why are categories adjacent? *Journal of Linguistics* 37:1-34.

———. 2004. *Efficiency and Complexity in Grammars*. Oxford: Oxford University Press.

Jaeger, T. Florian, David Orr, and Thomas Wasow. 2005. Comparing and combining frequency-based and locality-based accounts of complexity. Paper read at 18th CUNY Conference on Sentence Processing, at Tuscon, AZ.

Jaeger, T. Florian, and Thomas Wasow. 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.

Jespersen, O. 1922. *Language*. New York: Henry Holt.

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

Nissim, Malvina, Shipra Dingare, Jean Carletta, and Mark Steedman. 2004. An annotation scheme for information status in dialogue. Paper read at LREC.

Prat-Sala, M., and H. P. Branigan. 2000. Discourse constraints on syntactic processing in language production: A cross-linguistic study in English and Spanish. *Journal of Memory and Language* 42 (2):168-82.

Quirk, Randolph. 1957. Relative clauses in educated spoken English. *English Studies* 38:97-109.

Race, D.S., and M.C. MacDonald. 2003. The use of "that" in the production and comprehension of object relative clauses. Paper read at 26th Annual Meeting of the Cognitive Science Society.

Szmrecsányi, Benedikt M. 2004. On Operationalizing Syntactic Complexity. Paper read at JADT 2004 : 7es Journ´ees internationales d'Analyse statistique des Donn´ees Textuelles, at Louvain.

Tagliamonte, Sali, Jennifer Smith, and Helen Lawrence. 2005. No taming the vernacular! Insights from the relatives in northern Britain. *Language Variation and Change* 17:75-112.

Temperley, David. 2003. Ambiguity avoidance in English relative clauses. *Language* 79 (3):464-84.

Tottie, Gunnel, ed. 1995. *The man Ø I love: an analysis of factors favouring zero relatives in written British and American English*. Edited by G. Melchers and B. Warren, *Studies in Anglistics*. Stockholm: Almqvist and Wiksell.

van Nice, Kathy Y., and Rainer Dietrich. 2003. Task-sensitivity of animacy effects: Evidence from German picture descriptions. *Linguistics* (5):825–49.

Walter, Mary Ann, and T. Florian Jaeger. 2005. Constraints on complementizer/relativizer drop: A strong lexical OCP effect of that. Paper read at Annual Meeting of the Chicago Linguistic Society, at Chicago, IL.

Warren, Tessa, and Edward Gibson. 2002. The influence of referential processing on sentence complexity. *Cognition* 85:79-112.

Wasow, Thomas. 2002. *Postverbal behavior*. Stanford, Calif.: CSLI Publications.

Wasow, Thomas, and T. Florian Jaeger. in progress. Lexical Variation in Relativizer Frequency. Paper read at Expecting the unexpected: Exceptions in Grammar Workshop at the 27th Annual Meeting of the German Linguistic Association, at Cologne.

Zaenen, Annie, Jean Carletta, Gregory Garretson, Joan Bresnan, Andrew Koontz-Garboden, Tatiana Nikitina, M. Catherine O'Connor, and Thomas Wasow. 2004. Animacy Encoding in English: why and how. Paper read at ACL Workshop on Discourse Annotation, at Barcelona, Spain.

T. Florian Jaeger
Margaret Jacks Hall, Bldg. 460
Ling. Dept., Stanford University
Stanford, CA 94305-2150

tiflo@stanford.edu

Tom Wasow
Margaret Jacks Hall, Bldg. 460
Ling. Dept., Stanford University
Stanford, CA 94305-2150

wasow@csli.stanford.edu