

Processing as a Source of Accessibility Effects on Variation

T. FLORIAN JAEGER and THOMAS 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 (here *that*) 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 to appear; 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, 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. Coppock, E. Gibson, P. Hofmeister, R. Katzir, D. Orr, and the BLS 31 audience for their feedback. Special thanks go to H. Clark, D. Jurafsky, R. Levy, N. Snider, S. Thompson, and S. Vasishth for insightful discussions and invaluable comments, as well as 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 for part 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 studies have shown that NSRC subjects higher on the accessibility scale in (2) correlate with lower relativizer likelihood (see Table 1 for a summary).

Table 1: Relativizer Frequency by Subject NP Type of NSRC

Subject expression:	Pronoun		Lexical NP		
	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 and Thompson (to appear)	34%	41%		55%	

We wanted to see whether the observed correlation could be due to processing (as suggested in Race and MacDonald 2003; Hawkins 2001, 2004) and, more specifically, whether there is evidence that *degrees* of an NSRC subject’s accessibility influence relativizer likelihood. For this purpose, we constructed a large database of NSRCs in spoken English. Earlier studies of correlations between NSRC subject properties and relativizer likelihood were either conducted on rather small data sets (Fox and Thompson to appear; Temperley 2003; Tottie 1995) or exclusively on written data (Race and MacDonald 2003; Temperley 2003). Given our interest, spontaneous speech is a better source of data, for two reasons. First, informal spoken language is less subject to prescriptive influences, a potentially confounding factor. Second, any processing effects will show up more clearly in naturally occurring spontaneous speech, since spoken language is subject to real time processing pressures.

We automatically extracted all 4,405 NSRCs from the Paraphrase Switchboard corpus of informal spoken English.² Of these, 698 (25%) are introduced by

² 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.

Processing as the Source of Accessibility Effects on Variation

a *wh*-relativizer. Of the remaining 3,707 NSRCs, 56.6% start with a *that* relativizer and 43.4% have no relativizer. Inspection revealed that many of the *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*.³

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.⁴

Table 2: Givenness of an NSRC’s Subject and Relativizer Frequency

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

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*).⁵ Table 3 summarizes the average relativizer frequency for each of the subject types: relativizer likelihood increases as the accessibility of the NSRC’s subject decreases. 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 support earlier findings (e.g. Prat-Sala and Branigan 2000:180) arguing that accessibility is not a binary but a gradient property.

³ 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.

⁴ Here, “given” refers to referents that have been explicitly mentioned in the preceding dialogue. The Paraphrase Switchboard annotation of givenness is more detailed (see Nissim et al. 2004). The small sample size of “not given” NSRC subjects made more detailed comparisons impossible.

⁵ Over 92% of all NSRC subjects in our database fall into one of the six groups. The remaining NSRC subjects fell into groups that were either too small (e.g. proper names) and/or too heterogeneous (e.g. quantified NPs) with regard to accessibility to include them in the test below. We briefly discuss the effect of some of the special cases (e.g. generics, mass nouns, and quantified NPs) at the end of Section 3.4. For the effect of *that*-initial subjects on relativizer omission, see Walter and Jaeger (2005). NSRCs with *that* pronoun subjects exhibit extremely low relativizer frequency (18.2%). Walter and Jaeger attribute this to the lexical Obligatory Contour Principle.

Table 3: NSRC’s Subject and Relativizer Frequency in the Switchboard

Subject expression	Total	Relativizer Frequency
1 st person pronoun	1,905	39.7%
2 nd person pronoun	571	42.9%
3 rd 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	$\chi^2(5)=21.8, p<0.001$

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)? Following Race and MacDonald (2003), we measured the length (in words) of all NSRC subjects.⁶ 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 than NSRCs with multi-word subjects (57.3%; $\chi^2<1.6, p>0.2, N=339$). We conclude that the accessibility effect observed above is independent of the grammatical weight of an NSRC’s subject. That is, the accessibility effect holds after controlling for subject length (Race and MacDonald (2003) find the same in their study of 1,340 NSRCs from the *Wall Street Journal*). We turn next to the question of *why* this should be the case.

2. Accessibility Effects Are Explained by Processing

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 (Jaeger, Orr, and Wasow 2005; Wasow and Jaeger 2005), and ambiguity avoidance (Temperley 2003). Example (3) 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 a relativizer. (4) illustrates the effect of intervening material.

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

⁶ 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 measures of an NSRC’s subject complexity.

Both hearer-oriented (Hawkins 2001, 2004; Temperley 2003) and speaker-oriented (Race and MacDonald 2003) processing accounts of relativizer variation have been suggested. Following Ferreira and Dell's (2000) account for complementizer omission, Race and MacDonald (2003) propose that inserting a relativizer buys time to plan the production of an NSRC. In Jaeger and Wasow (2005), we have provided additional evidence supporting an account along those lines. Here we elaborate on a comment Race and MacDonald (2003) make in passing: "production difficulty factors [...] can influence the accessibility of the embedded [NSRC] subject and therefore 'that' use" (ibid, p. 951; square brackets added).

Given ample evidence that speakers tend to minimize production effort (Zipf 1949; Lindblom 1990) as long as this does not interfere with other constraints on communication, we hypothesize that speakers omit a relativizer wherever no other factor (including grammar) favors it. The need of speakers to hold the floor is such a constraint. It prevents speakers from falling silent when they aren't ready to start pronouncing the next phrase (as evidenced by the use of fillers, e.g.). We propose that omitting a relativizer is only advantageous for speakers if it does not conflict with the need to hold the floor, that is, if the speaker is ready to continue with the NSRC's first phrase (the NSRC subject). This predicts that speakers omit relativizers whenever they have already finished planning a "large enough" chunk of the NSRC subject. Speakers may have different strategies as to what constitutes a "large enough" chunk (and strategies may differ between different situations, cf. van Nice and Dietrich 2003). In any case, keeping a relativizer is one option for getting more time before the retrieval of the head of an NSRC's subject NP. Alternatively, speakers can insert fillers or lengthen words preceding the subject's head. However, for the vast majority of the NSRCs (92.5% in our sample), the head of the NSRC subject *is* the first (and only) word in the NSRC subject, which makes relativizers the only alternative to fillers.⁷ This way of thinking about relativizer omission is subtly different from Race and MacDonald's. Rather than claiming that speakers *insert* a relativizer whenever it buys them planning time, we propose that they only omit it if doing so is more efficient. This captures that *having* a relativizer is the default (we haven't come across any examples that are unacceptable *with* a relativizer).⁸ Now, how does accessibility enter the picture?

The more complex the production of the NSRC subject is, the longer it will take to plan it. We propose that, *ceteris paribus*, the more accessible a referent is in the current discourse model, the faster a corresponding expression is produced. Evidence for this claim 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

⁷ Note that, while the reasoning presented above is production-oriented, our main point (the correlation between accessibility and relativizer likelihood) is also compatible with certain comprehension-oriented accounts (as pointed out to us by Roger Levy).

⁸ Furthermore, there is evidence that speakers do not use relativizers to *alleviate* production difficulty, but rather to *signal* that they are having production difficulties (Jaeger 2005).

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).⁹ Further evidence that more accessible expressions are easier to process comes from the *comprehension* of NSRCs. Warren and Gibson (2002) present a series of self-paced reading experiments showing that decreased accessibility of an NSRC's subject increases reading times on the NSRC's verb.

The integration of accessibility effects into an independently motivated processing account makes the account outlined above a desirable one. In sum, we claim that the effects observed in Section 1 are due to faster construal of accessible NSRC subject referents. If this is correct, then other properties known to influence the time it takes to fully plan an expression (here: the subject expression) should also affect relativizer variation. This potentially includes any factor influencing lexical retrieval (e.g. lemma or word form frequency). In ongoing work, we also investigate to what extent the accessibility effect of an NSRC subject can be reduced to the predictability of the subject expression (or the first word of it) given the word immediately preceding the NSRC (usually the head noun).¹⁰ Here, we limit ourselves to the discussion of one factor: the *inherent* accessibility (Prat-Sala and Branigan 2000) of NSRC subjects.

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). Next, we will discuss three such inherent properties: number, referentiality, and animacy.

3.1. Number

Referents of singular NPs are inherently more accessible than referents of plural NPs (we assume that the construction of multiple references is more complex than the construction of a single reference). 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 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 (5) (all examples are taken from the Switchboard corpus).

⁹ Ferreira (1994) presents evidence that accessibility effects on word order are (partly) mediated via thematic role assignment. Thus the claim about accessibility and word order made above should be taken to apply primarily to prototypical subject roles (i.e. agentive subjects).

¹⁰ We are grateful to Dan Jurafsky for insightful discussions about the relation between predictability and relativizer likelihood. We plan to address this issue in detail in future work.

- (5) 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%; $\chi^2=9.5$, $p<0.01$, $N=465$) 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 2). 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*, and *it*) are likewise problematic since a great many uses of *it* (43%) occur in the idiomatic string ... *the way it* ...:

- (6) ...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 (see Fox and Thompson to appear for similar observations and discussion thereof). All instances of examples like (6) are *annotated* as inanimates in the Paraphrase Switchboard, but since, for most of these cases, it is questionable whether *it* refers to anything at all, this creates a strong confound *against* an animacy effect. Further complicating matters, NSRCs with *it* subjects modify

¹¹ It may be that impersonal references are less accessible (in part) because they are usually not anaphoric (and therefore not given). A similar argument could be made for animacy (discussed in Section 3.3). Both referential and animate referents are probably more likely to be the topic (i.e. what is talked about), which would make such referents more likely to be salient/accessible in the discourse. Hence referentiality and animacy effects could (in part) be due to *derived* accessibility.

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. Light head nouns strongly favor NSRCs without a relativizer (Fox and Thompson to appear; for a predictability-based account of this and other effects, see Wasow and Jaeger 2005; Jaeger, Levy, and Wasow 2005). Hence, unless NSRC subject animacy is an extremely strong predictor of relativizer absence, we would expect *fewer* relativizers for NSRCs with *it* subjects than for NSRCs with *he/she* subjects. This is indeed the case ($\chi^2=30.2$, $p<0.001$). 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 subjects 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; $\chi^2>10$, $P_s<0.005$, $N=133$ 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 (ARCs), as in (7), are much less likely (25.2%) to have a relativizer than NSRCs with object gaps (ORCs; 76.5%). This difference is mostly due to the preponderance of semantically light head nouns (favoring relativizer omission) for ARCs (70.4% vs. 31.1% for ORCs; $\chi^2=412.7$, $p<0.001$).

- (7) a. ... any time [_{NSRC} money and votes are involved] ...
b. ... the way [_{NSRC} our state tax is here] ...

Crucially, ARCs are also far more likely to have an inanimate lexical subject (78.2% of all cases) than ORCs (only 19.9% of which have an inanimate lexical subject). Once this confound is controlled for, no animacy or anti-animacy effect remains (all $\chi^2_s<1.2$). Since this null effect may be due to the small sample size after controlling for all confounds, we leave this issue open for future research.

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, Gen-

nari 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. 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), our samples of generics, mass nouns, and quantified lexical NSRC subjects (all arguably conceptually complex) exhibit 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 (6), or an existential *there*, as in (8), almost never have a relativizer (e.g. only 27.3% of the 22 NSRC with an existential *there* have a relativizer).

(8) ... anytime [_{NSRC} there is a change in weather ...]

Assuming that non-referring expressions are less complex than referring expressions, the low relativizer frequency that is correlated with non-referring NSRC subject expressions, if confirmed on larger datasets, would provide further evidence for the processing-based account outlined in Section 2.

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), shows 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. The two 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; Ferreira 1994; van Nice and Dietrich 2003) offers a plausible explanation for this fact. *Ceteris paribus*, formulation of highly accessible subject 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 accessibility is controlled for (but see Race and MacDonald 2003, who found a weak but significant effect in their sample of written language). Another factor that could account for the accessibility effects, predictability, will be investigated in upcoming research.

Using a large database enabled examination of more subtle accessibility effects on relativizer variation than shown in previous research, 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 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.

References

- Aissen, Judith. 2003. Differential object marking: Iconicity vs. economy. *Natural Language and Linguistic Theory* 21:435-483.
- Ariel, Mira. 1990. *Accessing Noun Phrase Antecedents*. London: Routledge.
- . 2001. Accessibility theory: An overview. In T. Sanders et al., eds. *Text representation: Linguistic and psycholinguistic aspects*. Amsterdam: Benjamins.
- Arnold, Jennifer E. 1998. Reference form and discourse patterns. Stanford University, Dept. of Linguistics.

Processing as the Source of Accessibility Effects on Variation

- 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, Thomas 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. *Proceeding of the (KNAW) Academy Colloquium: Cognitive foundations of interpretation, Amsterdam*.
- Ferreira, Fernanda. 1994. Choice of passive voice is affected by verb type and animacy. *Journal of Memory and Language* 33(6):715-736.
- Fox, Barbara A., and Sandra A. Thompson. To appear. 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. Paper read at the 18th Annual CUNY Sentence Processing Conference, Tuscon, AZ.
- Givón, Talmy. 1983. *Topic Continuity in Discourse: A Quantitative Cross-Language study*. Amsterdam: 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 UP.
- Jaeger, T. Florian. 2005. Optional *that* indicates production difficulties: Evidence from disfluencies. *Proceedings of Disfluency in Spontaneous Speech Workshop (DiSS'05), Aix-en-Provence, France*, 103-109.
- Jaeger, T. Florian, Roger Levy, and Thomas Wasow. 2005. The absence of "that" is predictable if a relative clause is predictable. Paper read at Architecture and Mechanisms of Language Processing, Ghent, Belgium.
- Jaeger, T. Florian, David Orr, and Thomas Wasow. 2005. Comparing and combining frequency-based and locality-based accounts of complexity. Paper read at the 18th CUNY Conference on Sentence Processing, Tuscon, AZ.
- Jaeger, T. Florian, and Thomas Wasow. 2005. Production complexity driven variation: The case of relativizer distribution in non-subject-extracted relative clauses. Paper read at the 18th Annual CUNY Sentence Processing Conference, Tuscon, AZ.
- Lindblom, B. 1990. Explaining phonetic variation: A sketch of the H&H theory. In W.J. Hardcastle and A. Marchal, eds., *Speech Production and Speech Modelling*. Amsterdam: Kluwer.
- Marcus, Mitchell P., Beatrice Santorini, Mary Ann Marcinkiewicz, and A. Taylor. *Treebank III*. Linguistic Data Consortium 1999.

T. Florian Jaeger and Thomas Wasow

- Nissim, Malvina, Shipra Dingare, Jean Carletta, and Mark Steedman. 2004. An annotation scheme for information status in dialogue. *Proceedings of the 4th International Conference on Language Resources and Evaluation (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):168-182.
- Quirk, Randolph. 1957. Relative clauses in educated spoken English. *English Studies* 38:97-109.
- Race, David S., and Maryellen C. MacDonald. 2003. The use of “that” in the production and comprehension of object relative clauses. *Proceedings of the 26th Annual Meeting of the Cognitive Science Society, Boston, MA*.
- Szmrecsányi, Benedikt M. 2004. On Operationalizing Syntactic Complexity. Le poids des mots. *Proceedings of the 7th International Conference on Textual Data Statistical Analysis*, 1032-1039.
- 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):464-484.
- Tottie, Gunnel. 1995. The man \emptyset I love: An analysis of factors favouring zero relatives in written British and American English. In G. Melchers and B. Warren, eds., *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* 5:825-849.
- Walter, Mary Ann, and T. Florian Jaeger. 2005. Constraints on optional *that* omission: A strong lexical OCP effect. Paper read at CLS 41, Chicago.
- Warren, Tessa, and Edward Gibson. 2002. The influence of referential processing on sentence complexity. *Cognition* 85:79-112.
- Wasow, Thomas. 2002. *Postverbal Behavior*. Stanford, CA: CSLI Publications.
- Wasow, Thomas, and T. Florian Jaeger. 2005. Lexical variation in relativizer frequency. Expecting the unexpected: Exceptions in Grammar Workshop at the 27th Annual Meeting of the German Linguistic Association, 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. *Proceedings of the ACL Workshop on Discourse Annotation, Barcelona*.
- Zipf, G. K. 1949. *Human Behavior and the Principle of Least Effort*. Cambridge, MA: Addison-Wesley.

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

Thomas Wasow
Margaret Jacks Hall, Bldg. 460
Ling. Dept., Stanford University
Stanford, CA 94305-2150
wasow@csli.stanford.edu