Cross-linguistic Variation in a Processing Account: The Case of Multiple *Wh*-questions*

Inbal Arnon, Neal Snider, Philip Hofmeister, T. Florian Jaeger & Ivan A. Sag
Stanford University

0. Introduction

The distinction between acceptable and unacceptable sentences has long served as the empirical basis for theoretical linguistics. Traditionally, two kinds of explanations have been offered to account for perceived unacceptability. In the first, unacceptability reflects the violation of grammatical constraints. The second explanation attributes unacceptability to processing complexity: a construction is judged unacceptable because it is hard to process.[1] The unacceptability of nested constructions like (1) has been attributed to their extreme processing complexity (Gibson 2000; Miller & Chomsky 1963). The acceptability of this construction improves when the processing load is reduced by replacing the lexical NPs with pronouns, as in (2).

(1) The boy **the cat the dog** bit scratched started crying
(2) The correspondent **everyone I** met trusts is interviewing the president

Interestingly, accounts of unacceptability in terms of processing complexity have been rare. Recent work, though, has focused on the possible advantages of identifying processing complexity as the source of perceived unacceptability. Several studies have demonstrated processing effects on the acceptability of structures previously considered to be ungrammatical (Featherston 2005; Kluender 2004, inter alia): when the processing load is reduced, acceptability is judged to increase. Some island constructions, for example, display improved acceptability when non-syntactic factors are manipulated (Keller 1996; Kluender 1998, 2004). Despite having the same syntactic structure, sentence (3), which has

* We wish to thank Igor Boguslavsky for access to the parsed Uppsala corpus, and Lev Blumenfeld, Tatiana Nikitina, and Dmitry Levinson for their patient help with the Russian corpus study. Thanks also to the 32nd BLS audience for a great Q&A session.
[1] We include under this definition accounts along the line of Hawkins 2004, in which processing preferences can be grammaticalized.

two lexical NPs, is judged as less acceptable than (4), which has two pronouns instead (Kluender 1998). Sentence (4), in fact, seems no different from (5), which is not an island construction:

(3) That was the play that **the teacher** wondered whether **a student** will like
(4) That was the play that **they** wondered whether **you** would like
(5) That was the play that they wondered if you would like

What are the advantages of a processing-based account of unacceptability? Such an account captures, in the case of islands, the effect of the referential form of the NP (lexical vs. pronominal), which would not ordinarily be encoded in a grammar. In general, a processing-based explanation can not only model gradience in the data, which is hard to capture within a categorical theory of grammar, but also provide motivation for it.[2] Moreover, it does so by appealing to factors that are independently motivated by other aspects of sentence processing. One challenge for processing-driven accounts is apparently categorical cross-linguistic variation in the acceptability of certain structures. Extending Kluender 1998, we argue that a processing-based account of islands can accommodate such cross-linguistic variation, and also better capture gradient variation.

In this paper, we will look at so called superiority violations (SUVs) in multiple *wh*-questions (e.g. *What did who read?*) to demonstrate the benefits of using processing complexity to account for perceived unacceptability. In the first part, we summarize recent studies demonstrating gradient acceptability in multiple *wh*-questions in English that can be partially captured by processing preferences (Featherston 2005; Hofmeister et al. to appear). These preferences, such as the distance between the filler and the gap (Gibson 2000) and accessibility of intervening material (Warren & Gibson 2002), are active in other unbounded dependencies. We suggest that the unacceptability of SUVs is due to processing complexity. Then, we review results about the processing of multiple *wh*-questions in other languages that a) reveal ordering preferences where there were claimed to be none (German) and b) confirm the apparent lack of them in others (Russian). These findings establish a cross-linguistic variation that is more subtle than previously assumed and which, we argue, cannot be accommodated within a categorical division of languages into those with ordering constraints and those without. In the third part, we suggest an account of cross-linguistic variation in which general processing preferences interact with language specific features. More specifically, we argue that the availability of case marking as a cue to thematic assignment underlies differences in the processing (and acceptability) of multiple *wh*-questions in English, German, and Russian.

[2]Theories that allow gradient grammaticality such as Stochastic OT (Boersma & Hayes, 2003) can capture gradient acceptability and provide a formal way of implementing it. However, one advantage of processing-based explanations is that they *motivate* specific constraints that cause the gradience.

1. Processing Accounts for Multiple *Wh*-phrase Ordering in English

English multiple *wh*-questions supposedly exhibit categorical order constraints. While (6a) is grammatical, the unacceptability of (6b) was originally explained by positing a language-specific constraint on *wh*-constructions that prohibits extraction over a syntactically 'superior' *wh*-phrase (Superiority constraint, Chomsky 1973). More recent accounts explain the distinction in terms of general grammatical principles (The Minimal Link Condition, Chomsky 1995), but still maintain a categorical difference in grammaticality between (6a) and (6b) (Pesetsky 2000):

(6a)	Who read what?	grammatical
(6b)	What did who read?	supposedly ungrammatical

Several findings raise the need to re-examine the supposedly categorical contrast between (6a) and (6b). First, despite being labeled 'ungrammatical', sentences like (6b) can be found (albeit rarely) in natural speech (Sag et al. under review; Clifton, Fanselow & Frazier 2006). More importantly, it has long been noted that there are exceptions to the 'Superiority constraint' in English: despite having the same syntactic configuration, sentences with *which*-phrases do not exhibit the contrast. (7a) and (7b) have both been argued to be grammatical (Bolinger 1978; Pesetsky 2000):

(7a) Which student read which book?
(7b) Which book did which student read?

An additional challenge to categorical grammatical accounts comes from evidence showing that *wh*-phrase ordering is conditioned by non-syntactic factors. Recent studies (Featherston 2005; Hofmeister et al. to appear) have demonstrated that SUVs display a pattern of gradient acceptability that is affected by the type of *wh*-phrase (bare phrase vs. *which* phrase), and is different from the simplistic contrast indicated by (6a) and (6b). Taken together, these findings suggest that the perceived unacceptability of SUVs like (6b) is gradient in nature and therefore not adequately captured by categorical grammatical accounts.

An alternative explanation that can capture the gradience of the data is given by the WH-Processing hypothesis (Hofmeister et al. to appear). Under this account, the relative rarity of structures like '*what did who do*' stems from their extreme processing cost: given the choice between several <u>grammatical</u> *wh*-orders (e.g. (6a) vs. (6b)), speakers disprefer those which (given the context) are associated with a greater processing cost. If multiple *wh*-questions are merely especially taxing unbounded dependencies, then they should be affected by the same factors that affect the processing of other unbounded dependencies (relative clauses, clefts, etc.). Combined with existing theories of processing complexity (Gibson, 2000; Warren & Gibson, 2002), the WH-Processing hypothesis makes the following predictions for a *wh*-filler-gap dependency:

I. Gaps that are further from the filler are harder to process.
II. Less accessible fillers make the dependency harder to resolve.
III. Less accessible intervenors make the dependency harder to resolve.

For fillers, accessibility refers to the boost in activation associated with the *wh*-phrase (Gernsbacher 1989; for further discussion, see Ariel (2001:47, 68), who uses the term *future accessibility*). For intervenors, we assume that higher accessibility consumes fewer resources and therefore is less taxing during dependency-processing (Warren & Gibson 2002).[3] SUVs have a bigger distance between the filler and the gap than their non-SUV counterpart. SUVs with bare *wh*-phrases, which correlate with lower accessibility than *which*-phrases (Frazier & Clifton 2002), are complex on all three levels (I – III). SUVs with *which*-phrases still have a larger distance between the filler and the gap than their non-SUV counterpart, but they have a more accessible filler and intervenor, rendering them less complex than bare SUVs.

In a series of studies, we (Hofmeister et al. to appear) investigated the extent to which the variance in acceptability of multiple *wh*-questions can be accounted for in terms of processing complexity by manipulating the distance between the filler and the gap and the accessibility of the filler and intervenor.[4] The acceptability results were complemented by on-line measures of processing complexity collected using self-paced reading. The results demonstrate gradient acceptability of SUVs that is accounted for by the same processing factors that have been shown to influence other unbounded dependencies (I – III). To give an example, we briefly summarize our tests of prediction II and III. We used SUV examples like (8-11) in which the accessibility of the filler and the intervenor was manipulated:

(8)	Pat wondered what who read.	BareFiller-BareIntervenor
(9)	Pat wondered what which student read.	BareFiller-*Which*Intervenor
(10)	Pat wondered which book who read.	*Which*Intervenor-BareFiller
(11)	Pat wondered which book which student read.	*Which*Filler-*Which*Intervenor

The results for acceptability judgments and reading times are shown in Figure 1 and 2. Figure 1 shows the log-normalized acceptability judgments for SUVs in the different accessibility conditions (higher numbers correspond to higher acceptability). Figure 2 shows the mean residual reading times for the verb region in the different conditions. Importantly, higher numbers correspond to slower reading times and reflect more difficulty.[5]

[3] Note that this notion of accessibility builds on, but differs from, the use in e.g. Ariel 2001. See Hofmeister *et al.* to appear for further discussion of how to apply the notion accessibility to *wh*-phrases.
[4] We assume increased processing cost correlates with reduced acceptability judgments (Fanselow & Frisch to appear; Hofmeister et al. to appear).
[5] Negative residual reading times mean the region is read faster than expected given its length.

Figure 1: Acceptability Judgments

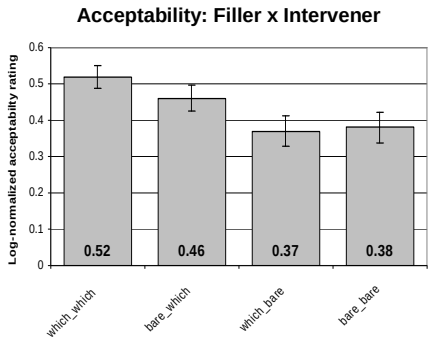
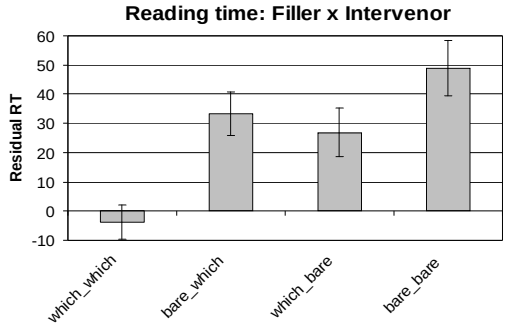


Figure 2: Residual reading times



SUVs with more accessible fillers and intervenors (11) show the highest acceptability ratings and fastest reading times. SUVs where only one of the entities was highly accessible (9,10) show lower acceptability and slower reading times, while SUVs with less accessible fillers and intervenors show the lowest acceptability ratings and slowest reading times (8).

To summarize, recent findings on multiple *wh*-questions in English reveal that a contrast historically regarded to be one of discrete grammaticality involves a gradient space of judgments that can partially be explained by independently motivated processing preferences. The gradience of the phenomenon and its sensitivity to processing factors is hard to accommodate within discrete models of grammar. Instead, the results support the role of processing complexity in creating the perceived unacceptability in English.

However, one apparent advantage of the grammatical account is its potential for accounting for cross-linguistic variation. A grammatical constraint in one language but not in another could explain the difference between languages that display ordering constraints (like English) and languages that don't (like German or Russian). How could one account for cross-linguistic variation within a processing account? If unacceptability is the result of general processing preferences, why don't we see the effect of these preferences across languages? Next, we review recent cross-linguistic findings that bear on the issue.

2. Cross-linguistic Variation

With regard to SUV constructions, the distinction between languages like English and languages like German or Russian is crucially based on the assumption that English, but not German or Russian, exhibits strict ordering preferences. The English data presented in the previous section demonstrates that the real space of judgments is more gradient than previously claimed and doesn't categorically rule out one of the orders. We now review additional work on multiple *wh*-questions in German and Russian that reevaluates the constraints on ordering in these languages.[6]

[6] Our work is influenced by previous work by Featherston (2005), but our interpretation differs.

The contrast between SUVs and non-SUVs has been claimed to be absent in German (Lutz 1996). Both (12a) and (12b) are considered equally acceptable.

(12a) Maria fragt wer was gelesen hat.
Maria asks who what read has
Maria asks who has read what

(12b) Maria fragt was wer gelesen hat.
Maria asks what who read has
Maria asks what who has read

Recent work, however, demonstrates the existence of ordering preferences in German that are consistent with the ones found in English. Featherston (2005) reports an ordering preference in German: SUV structures like (12b) were judged to be less acceptable than non-SUV structures like (12a). Moreover, the results demonstrate an effect of intervenor activation: SUVs with *which*-phrase intervenors (13) were judged more acceptable than ones with bare intervenors (14). No effect of filler activation was found:

(13) Was hat welcher Zahnarzt dem Patienten empfohlen?
what has which dentist to.the patient recommended
What has which dentist recommended to the patient?

(14) Was hat wer dem Patienten empfohlen?
what has who to.the patient recommended
What has who recommended to the patient?

Russian is also said not to exhibit a contrast between SUVs and non-SUVs: (15a) and (15b) are equally acceptable (Rudin 1988):

(15a) Elena staralas' razobrat'sia kto chto zakazal.
Elena tried figure-out who what ordered
Elena tried to figure out who ordered what.

(15b) Elena staralas' razobrat'sia chto kto zakazal.
Elena tried figure-out what who ordered
Elena tried to figure out what who ordered.

Indeed, Federenko (2005) found no ordering preference in Russian: both orders are judged equally acceptable and there is no effect of filler or intervenor accessibility.[7]

[7] It is important to note that Fedorenko (2005) confirms our results for English using the same methodology and almost identical stimuli as in the study on Russian multiple *wh*-questions.

To summarize, German exhibits ordering preferences similar to the ones in English, while no ordering preferences were reported for Russian. Importantly, a closer comparison of Featherston's (2005) results for German and our results for English suggests that the relative strength of these preferences differs between the languages. In both German and English, SUVs are judged as less acceptable than non-SUVs. But, the size of the 'Superiority effect' seems larger in English than in German. The empirical results from German and Russian reveal a cross-linguistic variation that is more subtle that previously claimed. Instead of a categorical difference between languages that have ordering constraints (e.g. English) and languages that don't (e.g. German and Russian), we see a more gradient distinction between languages that have strong, but not categorical, ordering preferences (e.g. English), languages that have weaker ordering preferences (e.g. German) and languages that don't seem to exhibit a preference (e.g. Russian).

3. Accommodating Cross-linguistic Variation in a Processing Account

There are several possible routes to account for cross-linguistic variation within a processing account. One is to say that the suggested preferences (distance is costly, less accessible intervenors are costly) are not universal: e.g. that in some languages, less accessible intervenors are not costly. This would require a bold statement about the directional influence of language on cognitive mechanisms, implying that different languages result in different processing mechanisms. Another is to say that the preferences are universal, but have different manifestations in different languages. For example, how distance is calculated can be different across languages. This is easy to apply for some preferences (e.g. distance), where language-specific features like free word order could affect the concept, but is harder to implement for other preferences (e.g. accessibility). A third route, and the one that we will argue for in this paper, is to say that processing preferences are universal, but that their strength depends on the availability of other cues.

Taking a step back, the motivating force behind the preferences we have identified as playing a role in the processing of *wh*-questions is to increase the ease of resolving an unbounded dependency. A bigger distance between the filler and the gap with more complex intervening material increases the difficulty of resolving the dependency. However, there are other language-specific features, like case marking or subject-verb agreement, which can assist resolution. In the absence of such additional cues, the 'violation' of word order preferences becomes more severe. For example, increased distance between the filler and the gap will induce more difficulty in the absence of additional cues that are used in resolving the dependency correctly (e.g. subject verb agreement, case marking, animacy preference for the Agent, etc.). How preferences surface in a given language will depend on the extent to which speakers can rely on other cues in on-line processing.

Next, we present an approach that draws heavily on the framework of the Competition Model (MacWhinney, 1987) but extends it to account for differences

in the surfacing of preferences in different languages. We propose an account of the reported cross-linguistic differences in *wh*-phrase ordering that attributes them to differences in the availability of case marking as a cue.

3.1. The General Approach

The idea that language specific features are responsible for cross-linguistic differences in on-line processing is not a new one. It is most fully elaborated in the Competition Model (MacWhinney 1987). In this framework, the relative strength of surface cues like word order, case marking and subject-verb agreement is responsible for differences between languages in on-line and off-line sentence interpretation. The strength of a cue is dependent on three factors. The first factor is availability, defined as the proportion of times that it is present. For case marking, availability would be calculated as the proportion of times that the noun has unambiguous case marking. The second factor is reliability, defined as the proportion of times where the cue marks the correct interpretation, when it is present. For case marking, this would mean the proportion of times a nominative-marked noun is the Agent of the sentence. The third factor is cost, which depends on the perceptual salience of the cue and the load it places on working memory.

The study of off-line thematic assignment demonstrates the different strength of surface cues in a variety of languages (MacWhinney & Bates 1989). The study of on-line thematic assignment reveals the effect of cue strength on reaction times. As predicted, the strongest cues lead to fastest reaction times and conflicting cues lead to inhibition and slowdown (Li, Bates & MacWhinney 1993). In an expansion of the model, Kempe & MacWhinney (1999) looked at the way that the availability of a cue is reflected in the processing benefits associated with it in on-line processing. Participants heard simple transitive sentences and had to identify the Agent as quickly as possible. While some sentences were ambiguous, others had various cues to the thematic assignment. The study manipulated the existence of cues like animacy, word order, and case marking. Kempe & MacWhinney report that more available cues, ones that are more frequent, had a bigger benefit in on-line processing. Looking at German and Russian they compared the availability and reliability of animacy, case marking, and word order using a corpus. While reliability is identical in the two languages (when case marked the thematic role of the noun is unambiguous), case marking is less available in German than in Russian: there are more ambiguously marked nouns in German. Accordingly, reaction times were more speeded when case marking was added in Russian in comparison to German. The results were interpreted as showing that because case marking is more available in Russian, Russian speakers rely on it more in on-line processing.

Drawing on these findings, we suggest that the cost of general processing preferences is mediated by the availability of other cues in the languages. In a language where case marking is a highly available cue, speakers will rely on that cue and will be able to ‘tolerate’ increased distance better. In other words, increased distance is costly across languages, as is lower accessibility, but the cost

of 'violating' those preferences is lower when other cues are highly available. In a fixed word order language, with no case marking cues, increased distance is very costly[8]. In a language with case marking, distance is increasingly less costly, depending on the availability of other cues. Can this kind of model explain the reported cross-linguistic differences in the processing of multiple *wh*-questions?

3.2. The Case for Case…

One striking difference between the three languages is the availability of case marking. Case marking on nouns is not an available cue in English. Case marking exists in both German and Russian, but the case marking paradigms of nouns in German are more ambiguous than the ones in Russian (Kempe & MacWhinney 1999 for declaratives). Crucially, this also seems to hold for the availability of case marking in questions words. The morphological paradigm of German question words is more ambiguous than that of Russian. Three out of the seven German question words are ambiguous between nominative and accusative case, while only three out of ten Russian questions words are.

We conducted a corpus study to test the hypothesis that German and Russian differ in the availability of case marking for question words. For German, we used the syntactically annotated TIGER (v. 1) and NEGRA corpora, which consist of 50,000 sentences (900,000 tokens) and 10,000 sentences (176,000 tokens) of newspaper text, respectively. For the Russian study, we used the dependency-parsed Uppsala corpus (Boguslavsky et al., 2002). The corpus consists of 17,772 sentences (256,034 tokens) of literary and informative text. We extracted the transitive embedded and non-embedded questions in each corpus that contained at least one *wh*-question word. This amounts to 167 questions in German, and 46 questions in Russian, with a total of 168 and 46 question words, respectively. To calculate the availability of case marking for question words, we calculated the percentage of question words that were unambiguously marked as nominative or accusative out of the total number of question words.[9] For German, only 11.3% of the question words in our sample are unambiguously case marked. In Russian, availability was three times higher: 34.8% of question words were unambiguously case marked. This difference is similar to the one reported in Kempe & MacWhinney for simple transitive sentences. The availability of case marking for nouns in their study was 68.5% for Russian and only 38.2% for German.[10] With the caveat that these two analyses were performed on different corpora, case

---

[8] In similar fashion, in this kind of language, the addition of case marking (on pronouns for example) may not be very beneficial since speakers are not used to relying on that information.

[9] For Russian we counted as unambiguous *kto*, *kogo*, *kakogo*, *kotorogo*, *kakaya*, *kotoraya*, *kakuyu*, *kotoruyu*, *kakikh*, and *kotorykh*, as well as *kakoj*, *kotoryj*, *kakie*, and *kotorye* when they were animate.

[10] Kempe & MacWhinney report the availability of case marking separately for the first and second NP in a sentence and separately for nominative and accusative case while we report the availability for a given question word, regardless of position or thematic assignment. The percentage we are using for the comparison was calculated by averaging the four measures.

marking seems to be more available for nouns than for question words. This could be due to the fact that the most common question word in both languages is the equivalent of English *what* (*was*, 71%; *chto*, 45.7%), which is not case marked in either language. Consistent with our hypothesis, we find that case marking in Russian *wh*-questions is more available than in German *wh*-questions.

The differing effect of the processing preferences could be attributed to the differential availability of case marking as a cue. The effect of distance and accessibility is masked when case marking is a highly available (Russian). The effect is apparent when case marking is less available (German) and is the strongest when case marking is not an available cue (English). The gradient cross-linguistic difference is thus attributed to the different availability of case marking.

3.3. Summary

The model we have presented views cross-linguistic variation as an interaction of processing preferences and language specific features. More specifically, we suggest that different costs are associated with dispreferred options, depending on the availability of other cues. The availability of information about the thematic assignment that is conveyed by cues like case marking or subject-verb agreement reduces the burden posed by variants with a high processing cost (e.g. those with increased distance and/or low accessibility intervenors). The 'cost' of dispreferred options is mediated by the strength of cues like word order, case marking, etc. that bias to the intended parse. With regard to unbounded dependencies, we suggest that increased distance and costly intervening material will affect the ease of resolving a dependency across languages, but the extent of the difficulty will depend on speaker's ability to draw on additional information. With regard to multiple *wh*-questions, we argue that the different manifestation of ordering preferences across English, German, and Russian can be partially attributed to the different availability of case marking in those languages. Case marking has been shown to be a more available cue in Russian than in German, both generally in transitive declaratives and more specifically in *wh*-questions; and case marking is more available in these languages than in English. This explains why the effect of distance and accessibility is most apparent in English, less so in German and not apparent in Russian.

4. Conclusions and Future Work

We have presented findings about the gradient nature of multiple *wh*-questions in English and offered a framework in which multiple *wh*-questions are seen as no different from other types of long distance dependencies. Their unacceptability is partially captured by processing factors known to affect other types of unbounded dependencies. We also sketched a proposal to account for cross-linguistic variation in ordering preferences by appealing to an interaction between general processing mechanism and language specific factors. Specifically, we argued that the different availability of case marking is responsible for the different surfacing of processing preferences in the languages tested (English, German, and Russian).

Why would one prefer the kind of account proposed here to one that uses categorical grammatical constraints to explain contrasts in acceptability within and between languages? First and importantly, the processing account can better capture gradient contrasts within languages and gradient differences between languages. As evidenced from the findings on multiple *wh*-questions, this kind of gradience is prevalent in language. Second, a processing account has wider coverage in explaining behavioral correlates of language use—the same mechanism that explains processing times is also assumed to underlie the relative acceptability of structures. Third, the processing account opens a fruitful research program into the ways that cross-linguistic variation can be modeled and quantified. For example, it makes the prediction that English multiple *wh*-dependencies should become easier to resolve when the likely thematic assignment is inferable from other cues, such as animacy of the *wh*-phrases. We are in the process of running experiments to test this hypothesis. We also predict that dependencies in case marking languages should become more difficult when the case marking cue is not available, as in the case for Russian multiple *wh*-questions with two inanimate entities. To conclude, we argue that an account that acknowledges the role of processing in multiple *wh*-phrase ordering will better account for variation in acceptability within and between languages.

References

Ariel, M. 2001. Accessibility Theory: an overview. In T. Sanders et al. In *Text representation: Linguistic and psycholinguistic aspects.* Amsterdam: Benjamins.

Boguslavsky, I., I. Chardin, S. Grigorjeva, N. Grigoriev, L. Iomdin, L. Kreidlin, N. Frid, 2002. Development of a dependency treebank for Russian and its possible applications in NLP. *Proceedings of LREC-2002*, Las Palmas. 3:852–856. http:/ruscorpora.ru/

Bolinger, D. L. 1978. Asking more than one thing at a time. In H. Hiz (Ed.), *Questions*. Dordrecht: D. Reidel 107-150.

Boersma, P. & B. Hayes. 2003. Empirical tests of the Gradual Learning Algorithm, *Linguistic Inquiry* 32: 45-86.

Chomsky, N. 1973. Conditions on transformations. In S. Anderson & P. Kiparsky (Ed.) *A Festricht for Morris Halle*. New York: Holt, Rimehart & Winston

Chomsky, N. 1995. *The Minimalist Program*. Cambridge, Mass.: MIT Press.

Clifton, C. J., Fanselow, G., & Frazier, L. 2006. Amnestying superiority violations: processing multiple questions. *Linguistic Inquiry* 37: 51-68.

Fanselow, G. & Frisch, S. to appear. Effects of processing difficulty on judgments of acceptability. In *Gradience in grammar*. G. Fanselow, C. Fery, M. Schlesewsky & R. Vogel (Eds.). Oxford: Oxford University Press.

Featherston S. 2005. Universals and grammaticality: *Wh*-constraints in German and English. *Linguistics* 43:4.

Fedorenko, E. 2005. Superiority violations in English and Russian: a cross-linguistic analysis. Presented at Ling-lunch talk series, MIT.

Frazier, L. & Clifton, C. 2002. Processing D-linked phrases, *Journal of Psycholinguistic Research* 31(6): 633-659.

Gibson, E. 2000. The Dependency Locality Theory: a distance-based theory of linguistic complexity. In Y. Miyashita, A. P. Marantz & W. O'Neil (Eds.), *Image, language, brain*. Cambridge, MA: MIT Press. 95–126.

Hofmeister, P., T.F. Jaeger, I. Arnon, I. A. Sag, N. Snider. to appear. Locality and accessibility in *wh*-questions. In *Linguistic Evidence 2006*: *Empirical, Theoretical, and Computational Perspectives*. Berlin: Mouton de Gruyter.

Keller, F. 1996. How do humans deal with ungrammatical input? experimental evidence and computational modeling. In *Natural Language Processing and Speech Technology: Results of the 3rd KONVENS Conference*. Dafydd Gibbon, (Ed.) Berlin: Mouton de Gruyter. 27-34.

Kempe, V. and MacWhinney, B. 1999. Processing of morphological and semantic cues in Russian and German. *Language and Cognitive Processes*. 14: 129-171

Kluender, R. 1998. On the distinction between strong and weak islands: A processing perspective. In *Syntax and Semantics 29: The limits of syntax*. P. Culicover & L. McNally (Eds.) San Diego, CA: Academic Press. 241-279.

Kluender, R. 2004. Are subject islands subject to a processing account? In *Proceedings of the 23rd WCCFL*. Vineeta Chand, Ann Kelleher, Angelo J. Rodríguez, and Benjamin Schmeiser (Eds.) Somerville, MA: Cascadilla Press.

Li, P., Bates, E. & MacWhinney, B. 1993. Processing a language without inflections: A reaction time study of sentence interpretation in Chinese. *Journal of Memory and Language* 32:169-192.

Lutz U. 1996 Some notes on extraction theory. In *On Extraction and Extraposition in German*. Lutz U. & Pafel J. (Eds). Linguistik Aktuell 11. Amsterdam: Benjamins.

MacWhinney, B. 1987. Toward a psycholinguistically plausible parser. In S. Thomason (Ed.) *ESCOL 1986*. Columbus, OH: Ohio State University.

MacWhinney, B. & Bates, E. (Eds). 1989. *The crosslinguistic study of sentence processing*. New York: Cambridge University Press.

Miller, G. A., & Chomsky, N. 1963. Finitary models of language users. In *Handbook of Mathematical Psychology*, vol. 2. Luce, D. R., Bush. R. R., & Galanter, E. (Eds.), New York: Wiley.

Pesetsky, D. 2000. *Phrasal Movement and its Kin*. Cambridge, MA: MIT Press.

Rudin, C. 1988. On multiple questions and multiple *wh*-fronting. *Natural Language and Linguistic Theory*. 6:445-502.

Sag, I. A., I. Arnon, B. Estigarribia, P. Hofmeister, T.F. Jaeger, J. Pettibone, and N. Snider. under review. Processing accounts for superiority effects. *Linguistic Inquiry*.

Warren, T. & Gibson, E. 2002. The influence of referential processing on sentence complexity. *Cognition* 85:79-112.