Superiority in English and German: Cross-language grammatical differences?

Jana Häussler, Margaret Grant, Gisbert Fanselow, Lyn Frazier

Abstract. Do the grammars of English and German contain a ban on moving the lower of two *wh*-s ('Superiority'), or is their lower acceptability due simply to the complexity of processing the longer dependency that results when the lower wh- is moved? The results of four acceptability judgment studies suggest that a processing-alone account is inadequate. Crossing *wh*-dependencies lower the acceptability of both German and English questions, but with a significantly larger penalty in English than in German (Experiment 1). The larger penalty in English cannot be attributed to greater sensitivity to violations in English, since relative clause island violations result in comparable effects in the two languages (Experiment 2). A processing-only account might claim long dependencies are easier to process in German than in English because of richer case, but a control experiment did not support this possibility (Experiment 4). We suggest that moving the lower of two *wh*-s is banned in the grammar in English but not in the grammar of German. This predicts that there should be a penalty for crossing dependencies in English even in helpful (Bolinger) contexts, confirmed in Experiment 3, and even in short easy-

[*] This work was supported by the German Research Foundation, grant no. FA 255/6-2. Parts of the study were presented presented at the Linguistic Evidence conference 2012 in Tübingen. We want to thank the audience for comments and discussion. We are especially grateful to Chuck Clifton, Sam Featherston and two anonymous reviewers for extensive comments on the manuscript. We thank Thomas Weskott for his contribution to develop the design of the study and the construction of materials, and Matt DuPont, Andreas Schmidt and Ümmühan Yildiri who were involved in data collection and analysis.

to-process sentences, confirmed by simple six word sentences in Clifton, Fanselow and Frazier (2006). Finally, if German grammar does not contain a ban on crossing, it is not surprising that the penalty in German is smaller than in English, or that like-Animacy of the two *wh*-s plays a larger role in German than in English since feature similarity generally gives rise to difficulty in processing whereas in English a grammatical ban on crossing will lower acceptability whether there is processing difficulty or not.

Keywords: Superiority, grammatical constraints, processing, grammaticizing

1 Introduction

Do all syntactic constraints a language respects belong to grammar, or can at least some of them be reinterpreted as effects of processing difficulty? This question has not only been raised recently with respect to established human parsing facts (e.g., Arnon et al. 2006, Hofmeister 2007, Hofmeister & Sag 2010, Kluender 1998), but it is also on the theoretical agenda of syntactic models such as the Minimalist Program (Chomsky 2008) which reduce the expressive power of grammar and must therefore shift the explanation of many generalizations to "third kind factors" of (language independent) processing.

The present paper contributes to this discussion by reporting several experiments concerned with the so-called superiority effect in English and German. It has been recognized since Kuno & Robinson (1972, also Chomsky 1973) that local movement (movement of the highest *wh*-element) is favored in multiple questions. Though (1a) is an ordinary un-noteworthy question of

English, (1b) is not. The same asymmetry is observed in embedded multiple *wh*-clauses as in (1c) and (1d). Movement of the lower *wh*-, as in (1b/d), is known as a Superiority violation.

(1) a. Who bought what?

b. What did who buy?

c. I wonder who bought what.

d. I wonder what who bought

Word order in multiple questions is particularly interesting for various reasons. First, while it was usually assumed in the syntactic literature that Superiority violations are ungrammatical due to the violation of some grammatical constraint (Chomsky 1973, 1981; Reinhart 1995; Pesetsky 2000), this view has recently been given up in Chomsky (2008).

Second, there appears to be considerable crosslinguistic variation. Superiority effects are very robust in English though they can be reduced or perhaps even cancelled by means of d-linking, cf. Pesetsky (1987), or context, cf. Bolinger (1978). In contrast, the syntactic literature on German typically presupposes that non-local movement in multiple questions does not lead to ungrammaticality (e.g., Haider 1993, Fanselow 2001) and attributes this to the overall grammatical organization of German which is claimed to render the constraints responsible for Superiority effects vacuous (but see, e.g., Wiltschko 1997). The available experimental evidence suggests that the crossing of a higher *wh* by a lower *wh* as in (2) leads to a weaker drop in acceptability (or sometimes none at all) in German than in English (cf. Featherston 2005, Fanselow, Schlesewsky, Vogel & Weskott 2011 and Fanselow 2010).

(2) a. Wer hat t_{wer} dem Patienten was empfohlen?

who has the.dat patient what recommended

`Who has recommended what to the patient?'

b. Was hat wer dem Patienten t_{was} empfohlen?

What has who the.dat patient recommended

In a recent study comparing Superiority effects in English and Russian, Fedorenko & Gibson (2008) presented multiple questions in contexts designed to insure a pair-list reading as in (3).

(3) Context: Helen works as a receptionist at a medical supply store. On Thursday, a large order of different devices was delivered to be distributed to fifty patients, but some paperwork was missing.

a. Helen tried to figure out who/which patient ordered what/which device.

b. Helen tried to figure out what/which device who/which patient ordered.

Fedorenko and Gibson observed a pronounced drop in acceptability for English sentences violating Superiority (3b) but no drop in acceptability at all for the corresponding Russian sentences[1]. Specifically, there were effects of word order and language as well as an interaction. All effects were due to a drop in acceptability of object-initial sentences in English (cf. (3b)),

[1] This finding confirms Rudin (1988) and others who claimed that Russian lacks a superiority effect. In a cross-linguistic study comparing Russian, Polish and Czech, Meyer (2004) however found a Superiority effect in Russian as compared to Czech and Polish.

especially for forms with an indefinite (*who*) subject. In Russian, there was no penalty for object-initial forms[2].

Third, a number of proposals have been advanced in the literature that try to derive superiority effects from assumptions concerning the processing of multiple questions. Based on cross-linguistic differences and the modulating effect of d-linking, Arnon et al. (2006) propose that Superiority effects are gradient and arise from the language processing system (see also Hofmeister et al. 2007). English, with its impoverished Case system, has the strongest Superiority effect, while an intermediate effect is found in German where Case is available but often ambiguous. Russian, where Case is available but ambiguous less often, exhibits no ordering preference in multiple *wh*-questions. Arnon et al.'s proposal is couched in the Competition Model of MacWhinney (MacWhinney, Bates & Kliegl 1984, Kempe & MacWhinney 1999) in which cue strength is determined by the availability, reliability and cost of using some cue. Cue availability is defined in terms of relative frequency (i.e. the proportion of times a particular cue is present) whereas cue reliability is defined as the proportion of times a cue unambiguously indicates the correct interpretation. Costs are determined by perceptual salience and memory load. The idea is that Case reliably leads to identification of OS structures in Russian, less reliably in German and not at all in English. In English Case marking is available only on

[2] Fedorenko and Gibson proposed that Superiority results from rigid word order in answers to questions in English. The required or desired parallelism between question and answer, they suggest, requires a SVO question because only an SVO answer is available. In Russian, scrambled orders (OVS) are available in the answer and therefore also in the question.

pronouns, though highly reliable when it is available. In contrast, Case marking is an available cue in both Russian and German but it is more often available in Russian. Arnon et al. provide two sources of supporting evidence for their model, experimental data of an agent identification task (Kempe & MacWhinney, 1999)[3] and corpus counts involving the Case ambiguity of *wh*-words, which both are problematic in our views[4].

Arnon et al. thus offer a parsing account of Superiority. Two sets of issues arise. One set involves evaluating their specific proposal for a processing account of crossover violations. The other set of issues concerns the general question of how, given a structure with degraded acceptability, one can distinguish a 'processing (only)' account, an account where the grammar does not contain a ban on the degraded structure, from a grammatically based account.

---

[3] Kempe & MacWhinney (1999) find that while Russian participants like German participants relied on Case marking in an agent identification task, Russian participants benefited more from Case marking in terms of processing speed. This relative advantage in Russian compared to German is actually restricted to accusative marking in OVS sentences. In all other cases, the two languages benefit equally from case marking.

[4] First, is it really appropriate to conclude that unambiguous Case marking is so highly available in Russian that it masks distance (Superiority) effects, while in fact the majority of instances of *wh*-words in the corpus are ambiguous? Second, it is not clear to us what the predictions are for multiple questions involving ambiguous *was* (‘what’) in German. Arnon et al. state “[...] 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.” This, however, refers to a situation in which both *wh*-phrases involved are ambiguous with respect to Case and hence the sentence is globally ambiguous whereas in the typical German example only one of the two *wh*-phrases, namely *was* (`what’), is ambiguous while the other one (*wer* `who’) is not.

In principle, given degraded acceptability of a structure in two different languages with possibly different degrees of degradation, various possible accounts are conceivable. Both languages might contain a grammatical prohibition against the structure, both languages might show processing difficulty only, or one language may contain a grammatical prohibition and the other exhibit only processing difficulty. In the present case, it has been proposed that both German and English contain a grammatical prohibition. Featherston (2005) suggested that Superiority-violations are ungrammatical in both languages, but they violate two constraints in English (a preference for SVO and a dispreference for in situ *wh*-subjects) whereas they violate only the latter in German.[5] The possibility that both languages show processing-only effects is precisely what Arnon et al. (2006) proposed. In what follows, we will compare that particular processing account with an account where the grammar of English contains an absolute ban on Superiority-violations whereas German only exhibits processing difficulty. Here and throughout we will distinguish between Arnon et al.'s specific processing account and other possible processing-only accounts by referring to the latter (processing accounts in general) as 'parsing accounts'.

Four acceptability rating experiments will be presented. They are designed to evaluate the Arnon et al. hypothesis and compare it to the hypothesis that in English, but not in German, there is a grammatical 'Superiority effect' (a grammatical constraint banning movement of the lower of

[5] Though there is indeed no constraint requiring SVO in German, there is nevertheless a SVO preference in terms of processing – evidenced in the context of ambiguity (for overviews see, Bader and Bayer 2006, Bornkessel

two *wh*-phrases, whether the constraint is specific to this configuration or an implication of some more general grammatical constraint). The structure of the paper is as follows. After discussing the Arnon et al. (2006) account in more detail in Section 2, in Section 3 we first outline the general aim and design of the study. This section then presents a multi-experiment study with precisely the same experimental and filler materials in German and in English. The sentences are presented to native speakers of the two languages for acceptability ratings. Section 4 takes up English and the evidence suggesting the existence of a grammatical Superiority effect in English; Section 5 turns to the impact of dependency length on the difficulty associated with superiority violations. Section 6 provides a summary of findings, arguments against the processing account in English, and evidence against a grammatical account in German.

Before turning to the details of the experiments, we must emphasize that several distinct issues run through this work. As noted above, we are centrally concerned with evaluating the processing-only account, as applied to both English and German, versus the account claiming Superiority violations lead to ungrammaticality in English but not German. Just as important, however, is determining what it means to say that a constraint is a grammatical constraint (or follows as one implication of a more general grammatical constraint). Although we delay full discussion of the issue until after the experiments, we will keep an eye on it as the discussion proceeds. Of course, grammatical constraints may have a processing related motivation in the first place (cf. Hawkins 1994, 2004). However, we assume that, by definition, for a constraint to

---

and Schlesewsky 2006) and in corpus data (e.g., Hoberg 1981, Kempen and Harbusch 2005, Bader and Häussler 2010a).

be a grammatical constraint entails being de-coupled from its source (if the constraint has a processing source). If a constraint is part of the grammar, then violating the constraint should give rise to degraded acceptability not only in configurations where it is motivated by processing-factors but in any instance where the constraint is violated. It is in particular in easy-to-process sentences that the status of a constraint will be clearest if, as we assume, processing complexity effects and ungrammaticality effects are related in a complex fashion rather than always being additive sources of complexity or always being interactive. Finally, determining whether there are cross-language differences in processing is important both to determining the correct interpretation of the data concerning our central hypothesis, but also in its own right.

2 A gradient processing account

Arnon, Snider, Hofmeister, Jaeger & Sag (2006) argue that Superiority is not a grammatical constraint but instead that the degraded status of sentences where a lower *wh*-phrase crosses over a higher one is due to the effects of various types of processing complexity:

(4) a. Gaps that are further from the filler are harder to process.

b. Less accessible fillers make the dependency harder to resolve.

c. Less accessible interveners make the dependency harder to resolve.

The first clause in (4) accounts for Superiority effects in general while (4b) and (4c) account for the modulating effect of the type of *wh*-phrase involved (e.g. *which*-phrase vs. bare *wh*-word). The processing account claims that it is not the crossing of the lower *wh* over the higher *wh* as

such that creates the effect but rather the resulting length of filler-gap dependency. Superiority-violating sentences have longer dependencies than sentences obeying Superiority, which might contribute to the difficulty of processing Superiority violations. Attributing the Superiority effect to length of dependency puts it on par with other structures involving shorter or longer dependencies, e.g., the well-studied processing difficulties associated with object-extracted relative clauses in comparison to subject-extracted relative clauses. It leaves open, however, why Superiority violations are perceived as ungrammatical while object-extracted relative clauses are only hard to process.

Under the hypothesis that bare *wh*- pronouns (*who, what*) are less accessible than d-linked *wh*-phrases (see Hofmeister, 2007), the amelioration of a Superiority violation with d-linked phrases might be attributed to (4b) and (4c) (though note that the status of the fronted object *wh*-phrase mattered only when the subject was d-linked in Fedorenko & Gibson's English study). In isolation, the claim in (4c) might appear somewhat counterintuitive. Thinking of interference, one would expect less accessible interveners to create less interference and thus make the dependency easier – not harder – to resolve. On the other hand, a less accessible intervener may induce a higher processing load leaving fewer resources for processing the dependency (e.g., in the logic of resource driven models like Dependency Locality Theory; cf. Gibson 2000). For instance, Warren & Gibson (2002) report less difficulties with object-extracted relative clauses when the relative clause subject is a first or second person pronoun and hence highly accessible in comparison to an indefinite NP. Similarly, a *wh*-subject in situ might be easier to cross if it is d-linked and hence accessible.

Taken together, three factors determine the ease to process a multiple *wh*-question: the length of the filler-gap dependency, the accessibility of the extracted *wh*-item, and the accessibility of the *wh*-item in situ. A fourth factor that the gradient processing account includes concerns Case marking. As noted above, the Arnon et al. embed their account in the framework of the Competition Model. We delay a more detailed discussion of that framework until later, but note only that it predicts cross-linguistic differences related to cue availability, how often a cue is present, and cue reliability, the trustworthiness of the cue when it is present. In particular, it predicts that word order 'cues' should be more important in English than German, but Case cues and animacy cues should be more important in German than in English (given that word order is more often ambiguous in German than in English).

3 Superiority and Animacy in yoked English and German studies

To assess the existence and size of a crossover penalty, i.e. a penalty for moving the lower *wh*- across the higher *wh*-, in English and German, we ran a joint study strictly comparable in the two languages. Although we think in general it is dangerous to interpret the absolute numbers in rating studies, or to compare the size of differences across studies since both are heavily influenced by experiment specific details such as the acceptability of the experimental sentences relative to the filler items, in the present study these factors were controlled to permit such comparison. Following the procedure suggested by Fedorenko and Gibson (2008) for comparison of English and Russian, relative clause island violations were included as a control (Experiment 2). Relative clauses are islands for extraction in both languages, and entirely

comparable in the two languages as far as we can tell. Therefore, they can serve as a kind of baseline and enable us to assess potential differences between subject populations. The study was a written acceptability rating study and contained several subexperiments including the two experiments described as Experiment 1 and Experiment 2 below, and Experiment 4 described in section 5. The experimental and filler items of English and German were translations of each other. Forty undergraduate students participated in the study in each language (at University of Massachusetts for the English part and University of Potsdam for the German part), for a total of 80 subjects.

Subjects were instructed to rate sentences on a scale from 1 (Totally Unacceptable) to 7 (Perfectly Acceptable). For the English study, subjects viewed the sentences on a computer screen using Linger software[6] and indicated their responses with a button press. The English sentences had a new randomized order for each subject. For the German study, subjects were tested using a pencil-and-paper questionnaire. Twelve versions of the questionnaire were created each with a different order. The order within each questionnaire was pseudo-randomized excluding consecutive items from same subexperiment and two items of the same condition in a row (even if separated by items from another subexperiment).

[6] Linger was developed by Doug Rohde at MIT, cf. http://tedlab.mit.edu/~dr/Linger/

3.1 Experiment 1: Comparing superiority violations in English and German

3.1.1 Method

Twenty lexically-different items were constructed. Each had four versions, as illustrated in (5). (for the complete set, see Appendix A). All sentences contained two bare *wh*-words. Either the higher- or the lower *wh*-word was moved. We will refer to this factor as Order. In two forms, (5a) and (5c), the subject appeared before the object; in two, (5b) and (5d), the object appeared before the subject, creating a Superiority violation. The different versions of a sentence also varied in the Animacy of the *wh*-object, and therefore in the match or mismatch in animacy between the two *wh*-words. In two forms, (5a) and (5b), both the object and subject were animate, whereas in the other two forms the subject and object differed in their animacy – the subject was again animate but the object was inanimate, cf. (5c) and (5d). We included this factor because it influenced the size of the crossing movement effect in German in previous studies (cf. Fanselow et al. 2011).

(5) a. The employees knew exactly who criticized who at the office.

b. The employees knew exactly who who criticized at the office.

c. The employees knew exactly who criticized what at the office.

d. The employees knew exactly what who criticized at the office.

(6) a. Die Angestellten wussten genau, wer wen im Büro kritisierte.
The employees knew exactly who.nom who.acc in-the office criticized

b. Die Angestellten wussten genau, wer was im Büro kritisierte.
The employees knew exactly who.nom what in-the office criticized

c. Die Angestellten wussten genau, wen wer im Büro kritisierte.
The employees knew exactly who.acc who.nom in-the office criticized

d. Die Angestellten wussten genau, was wer im Büro kritisierte.
The employees knew exactly what who.nom in-the office criticized

The corresponding German items are shown in (6). Two properties of German *wh*-words must be noted. Most *wh*-pronouns have an additional interpretation, viz. as an indefinite pronoun. The two readings are distinguished by position (in the left periphery of a clause, only the question interpretation is possible for (unmodified) *wh*-pronouns) and by accentuation (in situ *wh*-words have an indefinite interpretation if and only if they are deaccented). In a written questionnaire study, the prosodic property is not available to disambiguate for either interpretation. Fanselow and Féry (2008) reported an experiment in which participants had to read aloud (potential) multiple questions with ambiguous in situ *wh*-words. In nearly all of the cases (86 out of 96), participants accented the in situ *wh*-word. This demonstrates a strong preference for the question reading of a *wh*-item in German. Furthermore, the results of written and oral presentation of multiple questions were qualitatively identical; see Fanselow et al. (2011).

A second property of German *wh*-words is related to animacy. English *who* and *what* are equally ambiguous with regard to Case. German *was* (`what') shares this ambiguity, but *wer* and *wen* (`who') do not. In other words, for German the factor Animacy is tied to ambiguity. We will come back to this issue in the discussion section (section 3.3).

The experimental items were distributed over four lists following a Latin Square Design. Each participant saw only one list and thus each item only in one of its conditions, but an equal number of items in each condition. Within each list, the items were intermixed with 73 items from other sub-experiments (including Experiment 2 and Experiment 4) and 11 additional fillers used in prior experiments (thus allowing some comparison with earlier experiments), provided as Appendix G. The order of items was pseudorandomized as explained above.

3.1.2 Results

Statistical analyses were run using R (R Development Core Team, 2009) and the package *ez* (Lawrence, 2011). For this and all further experiments, the results were analyzed with ANOVAs by subjects and by items as random effect factors, and (where applicable) Language as an additional between-subjects (but within-items) factor. In general in this paper, we will report statistical tests on the data including both languages. However, analyses were first carried out on individual language data; for Experiment 1, these analyses are reported in Appendix C. Analyses of Experiment 1 include Animacy and Order (Superiority) as fixed effect factors.

The mean ratings for each condition in each language are shown in Figure 1. For both languages, we found a drop in acceptability scores when the order of *wh*-phrases does not obey Superiority, as confirmed by a main effect of Order ($F_1(1,78) = 291$, $MSE = 1.23$, $p < .001$; $F_2(1,19) = 755$, $MSE = .24$, $p < .001$). However, ANOVAs also reveal a significant interaction involving Order, Animacy and Language ($F_1(1,78) = 9.54$, $MSE = .28$, $p < .01$; $F_2(1,19) = 8.16$, $MSE = .16$, $p < .05$) reflecting a cross-linguistic difference in the size of Superiority effect and in the impact of animacy. The penalty for violating Superiority is more pronounced in English; this

produced a significant Order by Language interaction ($F_1(1,78) = 47.8$, $MSE = 1.23$, $p < .001$; $F_2(1,19) = 137$, $MSE = .21$, $p < .001$). The interaction of Order and Animacy is only partially significant ($F_1(1.78) = 4.65$, $MSE = .28$, $p < .05$; $F_2(1,19) = 2.57$, $MSE = .25$, $p = .13$) and basically driven by a significant interaction of the two factors in the German data ($F_1(1,39) = 14.20$, $MSE = .27$, $p < .001$; $F_2(1,19) = 9.88$, $MSE = .19$, $p < .01$). In German, the Superiority effect is slightly stronger when both *wh*-phrases refer to animates. For English in contrast, the Superiority effect is independent of animacy; the Order by Animacy interaction is not significant (Fs < 1). This difference in the impact of Animacy produced also a marginally significant interaction of Language and Animacy ($F_1(1,78) = 3.79$, $MSE = .30$, $p = .06$; $F_2(1,19) = 3.50$, $MSE = .16$, $p = .08$). As a main effect Animacy was not significant (Fs < 1). Finally, there was also a significant main effect of Language ($F_1(1,78) = 47.6$, $MSE = 3.00$, $p < .001$, $F_2(1,19) = 216$, $MSE = .33$, $p < .001$).

Insert Figure 1 about here

3.2 Experiment 2: Relative-clause island violations (control experiment)

Following a suggestion by Fedorenko and Gibson (2008), the study included a control experiment that contained relative clause island violations. The experiment involves long *wh*-movement and varies the extraction site and thereby the legitimacy of the extraction, cf. (7).

(7) a. Thomas accused the student of copying the term paper.

b. Thomas accused the student who copied the term paper.

c. What did Thomas accuse the student of copying?

d. What did Thomas accuse the student who copied?

Extraction out of a complement clause as in (7c) is fine whereas extraction out of a relative clause as in (7d) results in an island violation (Ross 1967). Relative clause island violations are known to produce robust effects in both English and German. This makes them suitable as a benchmark for evaluating the cross-linguistic difference in the strength of Superiority effects that showed up in Experiment 1.

3.2.1 Method

The control experiment examined 16 items like (7) and German counterparts as in (8) (for the complete material see Appendix B). All sentences also appeared in four versions corresponding to the four conditions that result from crossing two factors: The factor Movement varied the occurrence of movement and thereby the resulting sentences type. (7a) and (7b) are declarative sentences, (7c) and (7d) are *wh*-questions involving long extraction out of the embedded clause. The factor Clause Type varied the embedded clause which was either a complement clause as in (7a) and (7c) or a relative clause as in (7b) and (7d). As a result, the factor also varies extraction site for *wh*-movement and therefore the legitimacy of the extraction.

(8) a. Thomas beschuldigte den Studenten die Hausarbeit abgeschrieben zu haben.
Thomas.nom accused the student.acc the term-paper.acc copied to have

b. Thomas verwarnte den Studenten, der die Hausarbeit abgeschrieben hatte.
Thomas.nom cautioned the student.acc who the term-paper.acc copied had

c. Was beschuldigte Thomas den Studenten abgeschrieben zu haben?
What accused Thomas.nom the student.acc copied to have

d. Was verwarnte Thomas den Studenten, der abgeschrieben hatte?
What cautioned Thomas.nom the student.acc who copied had

The items were distributed over four lists, randomized and combined with items from other subexperiments and fillers as described above for Experiment 1.

3.2.2 Results

Figure 2 shows the mean ratings in Experiment 2. Analyses of variance reveal significant main effects of Movement ($F_1(1,78) = 1308$, $MSE = .49$, $p < .001$; $F_2(1,15) = 766$, $MSE = .34$, $p < .001$) and Clause Type ($F_1(1,78) = 743$, $MSE = .44$, $p < .001$; $F_2(1,15) = 214$, $MSE = .61$, $p < .001$) as well as a significant interaction of these two factors ($F_1(1,78) = 538$, $MSE = .50$, $p < .001$; $F_2(1,15) = 408$, $MSE = .26$, $p < .001$). Language failed significance as a main effect ($F_1(1,78) = 3$, $MSE = 1.08$, $p = .09$; $F_2(1,15) = 3$, $MSE = .42$, $p = .10$) but was fully significant in interaction with Movement ($F_1(1,78) = 47$, $MSE = .49$, $p < .001$; $F_2(1,15) = 44$, $MSE = .21$, $p < .001$) and in the interaction with the factor Clause Type ($F_1(1,78) = 31$, $MSE = .44$, $p < .001$; $F_2(1,15) = 14$, $MSE = .38$, $p < .01$). The interaction involving all three factors (Movement, Clause Type, and Language) was not significant ($F_1(1,78) = 3$, $MSE = .50$, $p = .07$; $F_2(1,15) = 2.26$, $MSE = .29$, $p = .15$).

Insert Figure 2 about here

As expected, relative clause island violations received low acceptability ratings. The penalty for extraction out of an island is basically the same in the two languages.

In German, but not in English, extraction out of a non-finite complement clause was penalized as well but to a lesser extent. The literature on extraction out of sentential arguments reports an asymmetry between finite and non-finite clauses with extraction out of the former being harder. An asymmetry between sentential subjects and sentential objects is also reported with extraction out of subjects being harder, for most speakers ungrammatical (for experimental evidence, see Jurka 2010). Extraction out of a non-finite complement clause as in (8c) should be the easiest case. Yet, there might be parsing problems in (8c). The parser may initially interpret the extracted *was* as a main clause argument in (8c). This is true for some of the English items as well but not to the same extent as for the German items.[7] Furthermore, having a main verb in the second

[7] In German, the extracted *was* could be the matrix subject (possible in English too but disambiguated almost immediately), the matrix object (temporarily possible for some of the English items) or a main clause adjunct (asking for a purpose or reason, roughly corresponding to English *why* but with the additional meaning of surprise or disapproval on the part of the speaker). Moreover, as Sam Featherston pointed out to us, some of the German items were globally ambiguous in that respect. Eliminating these items from the analysis increased the mean ratings but only slightly so and across the board, i.e. even in conditions not involving the potential ambiguity (6.61 for declarative sentences containing a complement clause, 5.14 for corresponding questions in which the *wh*- is extracted; 6.78 for sentences containing a relative clause, and 1.87 for relative-clause island violations; all effects are still fully significant).

position rather than an auxiliary may have contributed to the penalty for extraction out of a non-finite complement clause (cf. Grewendorf 1989, Haider 1993, Jurka 2010). We must leave the exact source for the extraction penalty to future work. For the moment, we want to emphasize the clear difference between extraction out of a complement clause and extraction out of a relative clause. For the latter, an island violation, we see the same strong penalty in German that we see in English.

3.3 Discussion of Experiments 1 and 2

The central finding of Experiment 1 is that the penalty for moving a lower *wh*-word over a higher *wh*-word was substantially larger in English than in German. This contrast in the size of the penalty for violating Superiority was significant as shown by the analysis involving Language as a factor. This finding is expected if moving the lower *wh*-word over a higher one is ungrammatical in English, but not in German, where the relatively small penalty may indicate a processing cost due to using a more complicated structure (one with a longer dependency) than is warranted. We delay until the General Discussion the question of whether the Arnon et al. (2006) processing proposal could provide a satisfying account of the data. Under the grammar-only view one could try to relate the difference between English and German to other differences between the languages with respect to extraction. Note, however, that overall English is more liberal than German while we observe the opposite when it comes to extraction of a *wh*-item across a higher *wh*-item, i.e. Superiority.

In German, but not in English, Animacy interacted with Order ('Superiority'): in the Superiority-violating form, sentences containing two animates were given lower acceptability

ratings than sentences with one inanimate.[8] Since *was* (`what') but not *wen* (`who.Acc') is Case ambiguous, one might wonder whether the interaction is due to the Case ambiguity of *was* rather than being due to its animacy. However, attributing the interaction to Case ambiguity would predict an effect of ambiguous *was* decreasing acceptability whereas the actual effect of *was* was to increase acceptability in the Superiority-violating questions. We consider it unlikely that our German participants initially mistook *was* as the subject and failed to reanalyze Case assignment after encountering the unambiguous *wh*-subject *wer*. There are two reasons why we exclude this scenario: Disambiguation by Case has been shown to be very effective (cf. Meng & Bader 2000);

---

[8] To address the role of animacy, the study also included sentences like (i). For German, neither animacy nor order had an effect; there was also no interaction. By contrast, the English translations of (i) showed a penalty for object-extracted relative clauses which we interpret as an effect of dependency length. This effect was modulated by animacy as reported in the literature (Traxler et al. 2002; Mak et al. 2002, 2006; Gennari & MacDonald 2009).

(i) a. Thomas berichtete uns von einem Regisseur, der einen Schauspieler rückhaltlos lobte.
T. told us about a director who a actor wholeheartedly praised

b. Thomas berichtete uns von einem Schauspieler, den ein Regisseur rückhaltlos lobte.
T. told us about a actor who a director wholeheartedly praised

c. Thomas berichtete uns von einem Regisseur, der einen Drehbuch rückhaltlos lobte.
T. told us about a director who a script wholeheartedly praised

d. Thomas berichtete uns von einem Drehbuch, das ein Regisseur rückhaltlos lobte.
T. told us about a script that a director wholeheartedly praised

furthermore our experimental material almost exclusively contains agentive verbs not compatible with an inanimate subject (cf. Appendix A) – hence, even if participants occasionally mis-parsed sentences as described above, the resulting sentences are degraded.

In Experiment 1, even the questions with unlike animacy showed a small crossover penalty in German. This comes as a surprise, given other experiments on German root clauses showing a (mild) crossover penalty only when the two *wh*-words have the same value for animacy (Fanselow et al. 2011, but see exp. 7 of Fanselow 2010). In Experiment 1, the question words appeared in embedded clauses. This was done to side-step the necessity in English to introduce an auxiliary for the questions involving a fronted object. The discrepancy between German root clauses, where there is a penalty only for like animacy, and German embedded clauses, where there is a small crossover penalty also when the subject *wh*-word is animate and the object *wh*-word is inanimate, may indicate that especially in root clauses there may be an information structure motivation for fronting the lower of two *wh*-words. However, we suspect that the difference between root and embedded clauses may not be trustworthy.[9]

The relative clause island violation penalty was just as severe in German as in English, removing the possibility that German participants are simply less willing to give low acceptability ratings. Interestingly, declarative sentences containing a relative clause (*Thomas accused the student who copied the term paper*) were rated lower in English than in German. We don't know why this is.

[9] A pilot study on this matter exhibits comparable effects in root questions and embedded questions.

It has been proposed that relative clause islands may also largely reflect processing difficulties – indeed the same difficulties offered as an explanation for Superiority (Hofmeister & Sag 2010; for a critical review of the arguments for reductionist accounts see Phillips 2013). The data from Experiments 1 and 2 show that it is not possible to maintain both the proposal that Superiority is due to the processing constraint of Arnon et al., which is based on dependency length, and the claim that the penalty for extracting from relative clause islands is largely due to processing. In this case, there would be no explanation for German showing a relative clause island violation penalty as large as in English, but a Superiority violation penalty that is substantially smaller than in English. If a stronger Case cue makes non-local dependencies easier in one structure, it should also make them easier in the other.

4 Superiority is a grammatical condition in English

There are many indications that Superiority is a grammatical condition in English and not simply the result of piling up various sources of processing complexity. If processing complexity were the (only) source of the degraded acceptability of Superiority violations in English, we would expect that in short, simple-to-process sentences there would be little or no indication of degraded acceptability when the lower *wh*-word is fronted. But this is not the case. For example, in Clifton, Fanselow & Frazier (2006) there was a large penalty for Superiority violations even in very simple six word sentences. We think this is hard to reconcile with an account that reduces Superiority effects to processing difficulties.

In this section, we present the results of another experiment that further challenges the processing hypothesis and supports the assumption that Superiority is a grammatical constraint in English. Experiment 3 exploits sentences like (9) introduced by Bolinger (1978) and claimed to lack Superiority effects.

(9) I know who was supposed to do what, but what did who actually do?

Bolinger sentences can be characterized by the presence of context already introducing paired *wh*-words with contrastive focus placed on the item following the second *wh*-. In (9), *supposed to* triggers a non-actuality implicature. Non-actuality implicatures are present in many Bolinger examples and known to aid processing in the case of ellipsis (see Grant et al. 2012).

4.1 Experiment 3: Bolinger sentences

According to one possible parsing account of Superiority violations that penalizes the use of an unnecessarily complex question without any reason, information structure might be expected to eliminate whatever oddness results from using a complex structure. When information structure warrants the use of the more complex question, one with movement of the lower *wh*-word, the Superiority violation may be acceptable. Indeed, Bolinger (1978) argued that Superiority violations are acceptable in English when judged in the right context such as (9). In the literature, his judgments have often been taken for granted, and considered a mystery for a grammatical account of Superiority violations.

The sentences in Experiment 3 consist of two conjuncts concatenated by the conjunction *but*. The first conjunct contains a matrix clause and an embedded clause which introduces a multiple

wh-question. This question occurs either directly after the matrix verb or further embedded under a non-actuality triggering verb. The second conjunct comments on the question by posing a related question. The presence of such a context may aid acceptability for several distinct kinds of reasons. It may be the availability of any context per se. It might be the fact that the context makes the content of the question already familiar. It might be the introduction of an implicit Question-under-Discussion (e.g., the actuality of the state of affairs described) that facilitates comprehension or acceptability of the second clause. Alternatively, facilitation might result from the availability of a sharp contrast between the context and the multiple question. Of course, these potential sources of facilitation are not mutually exclusive. In any case, the question behind Experiment 3 was whether choice of the more complicated question with the longer dependency is fully acceptable when motivated by a Bolinger context or whether a penalty is still observed when the lower *wh*-word is moved, as expected if Superiority is a grammatical constraint.

4.1.1 Method

24 sentences containing multiple questions were constructed (see Appendix E). Half of them contained a non-actuality triggering verb like *was supposed to* in the first conjunct, cf.(10), (see Grant et al. 2012, for discussion of how such modals introduce an implicit Question-under-Discussion). The other half of the sentences did not contain a non-actuality triggering verb. The non-actuality implicature (NAI) was reinforced by one of the adverbials *actually* or *in reality* in the second conjunct. Sentences with no non-actuality implicature in the first conjunct contained some other clear contrast driven by antonymic adverbs in the two *wh*-clauses, e.g. *legally-illegally* as in(11), which set up a clear contrast but not one based on actuality/non-actuality.

Later, two items of the (10) type were excluded from analyses due to a coding error. In two conditions ((10a) and (11a)), the lower *wh*-word is fronted; in the other two ((10b) and (11b)) the higher *wh*-word is fronted.

(10) NAI-inducing element

a. I know who was supposed to do what … but what did who actually do?

b. I know who was supposed to do what … but who actually did what?

(11) No NAI-inducing element

a. I know who legally did what … but what did who illegally do?

b. I know who legally did what … but who illegally did what?

Forty-eight subjects rated the naturalness of the sentences on a scale ranging from 1 (totally unnatural) to 5 (perfectly natural). Subjects were tested individually on computer terminals using Linger software. This experiment was run only in English because the question is whether a grammatical Superiority violation penalty is still present in Bolinger-type contexts. The use of a 5-point scale does not reflect a theoretically-driven decision, but only the standard practice of the laboratory where the study was conducted.

4.1.2 Results

The results of Experiment 3 were analyzed using ANOVA with Order (and therefore the absence/ presence of a Superiority violation) as a within-subject and within-items factor and Context (the presence or absence of an NAI) as within-subjects, but between-items factor. As mentioned in the

method section only 22 sentences entered the analyses due to coding errors. Figure 3 shows the means for each condition.

Insert Figure 3 about here

Analyses of variance showed that a fully significant Superiority violation penalty is observed when the lower *wh*-word is moved ($F_1(1,47) = 44.27$, $MSE = 1.68$, $p < .001$; $F_2(1,20) = 81.78$, $MSE = .20$, $p < .001$). An advantage in ratings for Non-Actuality Implicature contexts over contexts establishing some other contrast was significant by subjects but not by items ($F_1(1,47) = 4.68$, $MSE = .89$, $p < .05$; $F_2(1,20) = 2.75$, $MSE = .39$, $p = .11$). The interaction between Order and Context failed significance ($F_1(1,47) = 2.20$, $MSE = .66$, $p = .14$; $F_2(1,20) < 1$, $MSE = .20$, $p = .36$).

4.1.3 Discussion

Experiment 3 attests to the existence of a Superiority effect even in sentences providing a context facilitating processing by providing a contrast and some content in advance. This finding is consistent with the conclusion that a grammatical constraint bans Superiority violations in English. Although the results do not exclude a parsing account in general, they at least cast doubt on a version of a parsing account under which one would expect the longer dependency associated with the Superiority violation to be acceptable when supported by context. The superiority effect in Experiment 3 is independent of the presence or absence of a non-actuality implicature. This is compatible with the grammar account but also with a processing account. The non-actuality is just one way among others to motivate the repetition; the other contrast used

in the experiment was only slightly less effective in this regard (or may just less natural) as witnessed by the main effect of Context.

Bolinger contexts are comparable with effects of d-linking (Pesetsky 1987), as two reviewers pointed out to us. D-linked Superiority-violations have been noted to be much more acceptable than violations involving indefinite *wh*-constituents. Whether this is due to the availability of an implied context or whether it is due to a grammatical distinction among types of interrogative phrases is an interesting and unresolved question. Hofmeister et al. (2007) and Hofmeister et al. (2013) presented convincing evidence that the accessibility of a constituent during memory retrieval is determined in part by its complexity. This may be part or possibly all of the reason why Superiority-violations are at least ameliorated with d-linked phrases. However, determining whether d-linking can completely eliminate a Superiority-violation would require experimental investigation, we think, and it goes beyond the scope of the current paper.

4.2 Parallelism

A preference for parallel structures might offer a different sort of context motivating Superiority-violations. This possibility was investigated in a separate experiment (to be reported elsewhere) testing examples like (12) and (13) where the multiple question in the final clause did (a, c), or did not (b, d), have an order parallel to that of the first clause, which contained a d-linked interrogative.

(12) a. Which gallery owner considered which young painters as exceptionally talented, and who therefore promoted who?

b. Which young painters did which gallery owner consider as exceptionally talented, and who therefore promoted who?

c. Which young painters did which gallery owner consider as exceptionally talented, and who did who therefore promote?

d. Which gallery owner considered which young painters as exceptionally talented, and who did who therefore promote?

(13) a. Welcher Galeriebesitzer betrachtete welche jungen Maler als besonders
Which gallery-owner considered which young painters as particularly
vielversprechend und wer hat wen daher gefördert?
promising and who.nom has who.acc therefore promoted

b. Welche jungen Maler betrachtete welcher Galeriebesitzer als besonders
Which young painters considered which gallery-owner as particularly
vielversprechend und wer hat wen daher gefördert?
promising and who.nom has who.acc therefore promoted

c. Welche jungen Maler betrachtete welcher Galeriebesitzer als besonders
Which young painters considered which gallery-owner as particularly
vielversprechend und wen hat wer daher gefördert?
promising and who.acc has who.nom therefore promoted

d. Welcher Galeriebesitzer betrachtete welche jungen Maler als besonders
Which gallery-owner considered which young painters as particularly
vielversprechend und wen hat wer daher gefördert?
promising and who.acc has who.nom therefore promoted

Parallelism did not eliminate the crossing violation in English or in German (where a significant penalty for crossing was observed, presumably because we tested questions with two *wh*-words of like-animacy). Indeed, the effect of parallelism was not significant in either language, despite having comparable statistical power to the studies described here (and a fully significant relative clause island violation effect for the same control sentences tested in Experiment 2).

Perhaps there is some context that would eliminate the penalty for moving a lower *wh*- over a higher *wh*-. What we have shown here is simply that a Bolinger context is not sufficient, nor is a context containing parallel structure.

5 The impact of dependency length differences

If superiority effects are indeed due to the length of the filler-gap dependency as claimed by Arnon et al.'s gradient processing account, we should see the same asymmetry between *wh*-subject extraction and object extraction in single *wh*-questions, i.e. independent of crossing.

Object questions should be harder to process than subject questions. In principle, this should hold for English and German likewise.[10] Furthermore, the Arnon et al. account predicts that case marking helps in German but not in English; the penalty for object extraction should therefore be stronger in English.

To test this prediction, Experiment 4 examines sentences similar to those in our Superiority study (Experiment 1 of the present paper) but involving only one *wh*-dependency. Experiment 4 was run as a subexperiment together with Experiment 1 to allow direct comparison with that study.

5.1 Method

Twelve item sets like (14) and (15) were created for each language (for the complete list see Appendix F). Each set has one version with extraction from subject position (a), and one with extraction from object position (b). Experiment 4 was part of the multi-experiment study which also contained Experiments 1 and 2; thus the subjects and procedure were the same as for those experiments.

(14) a. The bartender noticed who hid something during the police raid.

b. The bartender noticed what someone hid during the police raid.

[10] If dependencies between arguments and the verb are also assessed, then German will show a larger penalty for each argument than English does, since in English but not German the arguments are adjacent to the verb. However, this predicts incorrectly that the German sentences in Experiment 1 should have been rated as less acceptable than their English counterparts. We will thus tentatively assume that wh-dependencies are the major determinants of complexity or acceptability in multiple questions.

(15) a. Der Barmann bemerkte, wer etwas während der Polizeirazzia versteckte.

The bartender noticed who something during the police-raid hid

b. Der Barmann bemerkte, was jemand während der Polizeirazzia versteckte.

The bartender noticed what someone during the police-raid hid

5.2 Results

The results of Experiment 4 are shown in Figure 4. Overall, the German participants gave slightly higher scores than the participants at the University of Massachusetts. This produced a main effect of Language ($F_1(1,78) = 7.61$, $MSE = 1.53$, $p < 0.01$; $F_2(1,11) = 10.34$, $MSE = .34$, $p < .01$). Extraction and thereby dependency length had only a minor effect and failed full significance ($F_1(1,78) = 4.90$, $MSE = .26$, $p < .05$; $F_2(1,11) = 1.54$, $MSE = .24$, $p = .24$). In English, sentences involving subject extraction and sentences involving object extraction received virtually identical scores (5.22 vs. 5.15). German participants, in contrast, rated for subject extraction somewhat higher than object extraction (5.87 vs. 5.58). However, there was no significant interaction between Language and Extraction ($F_1(1,78) = 2.05$, $MSE = .26$, $p = .16$; $F_2(1,11) < 1$, $MSE = .17$, $p = .35$). Analyses of the German data show also no fully significant effect of Extraction ($F_1(1,39) = 6.66$, $MSE = .26$, $p < .05$; $F_2(1,11) = 1.76$, $MSE = .29$, $p = .21$).

Insert Figure 4 about here

5.3 Discussion

Experiment 4 did not detect a dependency length effect in German or in English. If the experiment were run in isolation we would have to stop here. However, given that we found a

difference for subject-extraction versus object-extraction in Experiment 1 within the same study, we can go on and conclude that the effect in Experiment 1 must have to do with crossing rather than being due to differences in the dependency length alone. In other studies, English shows effects of dependency length, e.g., for relative clauses (cf. Gibson 2000 and references therein) and in German there is some variability. The lack of a dependency length effect in Experiment 4 might be related to the fact that the distance difference is rather small and the intervener is light (an indefinite pronoun). So, while we do not want to draw particularly strong or general conclusions for this study, we do want to emphasize that in the same study where we found a stronger effect of Superiority violations in English than in German, Experiment 1 of the present paper, we failed to find the same effect of dependency length in single questions. This makes it difficult to attribute the effects of Superiority violations to differences in the complexity of long dependencies in the two languages.

In summary of this section, the predictions that can be derived from the processing account proposed by Arnon and colleagues are not borne out, though they are not unambiguously disconfirmed either. Under the view that Superiority effects are related to the longer dependency and modulated by the strength of Case and animacy cues we expect stronger dependency length effects in English compared to German as well as a stronger impact of animacy in German. However, Experiment 4 revealed no general penalty for extracting the object compared to extracting the subject. In consequence, cross-linguistic differences regarding the relative strength of such a dependency length effect as well as the impact of animacy are hard to evaluate. Yet, the results do not show the pattern predicted by the processing account.

6 General Discussion

Superiority effects in English and German differ – both quantitatively and qualitatively. The size of the penalty for extracting the lower of two *wh*-words is substantially larger in English than in German. This is what would be expected if Superiority is due to a grammatical constraint in English but not in German. Further evidence for a grammatical constraint in English stems from the persistence of a large penalty for Superiority violations regardless of whether the sentences are extremely short and simple, as in Clifton et al. (2006), whether movement of the lower *wh*-word would be motivated by context (Experiment 3), or parallelism (cf. section 4.2). Though the focus of the current paper is on determining the status of Superiority violations, whatever the exact nature of the constraint is it must be a specific constraint since locality shows effects that are at least every bit as large in German as in English.

We turn now to the Arnon et al. processing account of Superiority, which is based on findings of weaker effects in languages with 'stronger' Case cues. The first point to note is that in English when Case does disambiguate the analysis of a sentence, it is fully effective. Traxler and Pickering (1996) show an immediate effect of Case when it disambiguates structure in English. The general point that 'cues' that disambiguate are equally strong across languages independent of how often the cue is present across sentences of the language is also mentioned by Fedorenko and Gibson (2008). The appearance of stronger cues for some languages has largely come from studies pitting word order in English against word order in some other language where a Noun-Verb-Noun sentence is *grammatically* ambiguous. The fact that animacy in English in Noun-Verb-Noun does not give rise to an OVS analysis but that it does in, say, Italian or German need

not show that word order is a stronger ‘cue’ in English than Italian or German, and animacy is a stronger ‘cue’ in Italian and German than in English. It might just show that 'cues' that disambiguate ambiguous sentences act as if they are stronger because the input is grammatically ambiguous and thus open to an alternative to the structurally preferred analysis.

Setting aside general worries about the Competition framework per se, the gradient processing account is problematic even internal to the data presented here. One motivation for this conclusion is the finding of no dependency length effect in Experiment 4. If at all, we saw a nonsignificant numerical trend suggesting a dependency length effect in the German data, not the English data, whereas the Cue competition account would have predicted a stronger effect in English than in German (especially since this was observed in precisely the same conditions where English shows a stronger Superiority-violation effect than German). The comparability of English and German with respect to relative clause island violations strengthens this point, we think, because the violations in the two languages are entirely comparable. Indeed, the results make it unconvincing to claim that island violations and Superiority are both cases of processing complexity alone, since the two languages differ with respect to Superiority but not with respect to relative clause island effects. Furthermore, we found no corresponding dependency length effects in sentences with a single *wh* (Experiment 4). Apparently, it is not the longer dependency as such which makes a sentence not obeying Superiority degraded in acceptability.

Perhaps a parsing or grammatical account could be based on penalizing unnecessarily-long dependencies, i.e., an economy principle, rather than a principle based on dependency length per se. First we note that such an account is unlikely to really be a parsing-only account since the

parsing literature does not support the idea that on-line parsing involves transderivational comparisons. However, the grammar might indeed contain a Superiority-constraint couched as an economy condition. Our arguments based on dependency-length alone would not address such an account. However, the interaction with animacy in German but not in English in Experiment 1 would still need to be explained on such an account.

Before leaving the processing account based on dependency length, we wish to note that the account would render other structures in English puzzling at best. In examples like (16) we might expect the longer dependency, between *which soup* and the object of *in,* to be excluded and only the shorter dependency allowed. Instead what we find is that only the longer dependency is available. This results in a silly interpretation in (16), involving pots cooking in soups.

(16) Which soup is this pot easy to cook __ in __?

a. It is easy to cook this pot in which soup.

b. *It is easy to cook which soup in this pot.

As noted by Fodor (1978, also Pesetsky 1982) it is only the nested dependency corresponding to (16a) which is grammatical in English, despite the fact that the nested pattern of dependencies requires a longer dependency than does the ungrammatical intersecting assignment of dependencies.

What does it mean to say that a constraint is 'in the grammar'? The understanding of this issue that is often put forward is that a grammatical constraint is categorical, not gradient (e.g.,

Hofmeister, Staum-Casasanto & Sag 2010). Though we think there is something to this claim, note that behavioral data—and of course linguistic judgments are one kind of behavior – always involve the performance system in addition to the grammar. So the real content of the claim that grammatical constraints should be categorical amounts, we suspect, to the claim that they should be detectable even in the simplest sentences, removing as many complicating factors as possible. Superiority violations in English pass this test: as already noted above, large Superiority violation penalties are observed in English in extremely simple sentences.

We have to admit that a Superiority violation penalty can be observed in German too for comparably simple sentences (cf. Featherston 2005, Fanselow et al. 2011). This might suggest that a grammatical factor is at play in German, too, which would also allow us to understand why German shows no parallelism effects (cf. section 4.2), just like English. The reduced size of the Superiority penalty might indicate that crossing movement violates fewer constraints in German than in English (Featherston 2005), or that the constraints banning crossing movement are "softer" in German, in the sense of Sorace and Keller (2005).

However, the degraded status of multiple questions with crossing movement is clearly visible for sentences only in which the two *wh*-s have like animacy. Multiple questions with crossing movement do occur in German corpora when the two *wh*-s differ in animacy (Featherston 2005), so they are clearly grammatical (but crossing is only half as frequent in multiple questions as in comparable declarative clauses, Fanselow et al. 2011), and the reduction of acceptability is smaller in this case (and can sometime not even be detected at all). In our view, the impact of animacy and more precisely of similarity in animacy points towards a parsing explanation in

German (cf. Fanselow et al. 2011).[11] As emphasized already, we see no corresponding impact of animacy in English. We take this as supporting evidence that (i) the observed effects have different sources in English and German, and (ii) we are dealing with a grammatical constraint in English.

If a constraint is contained in a grammar, presumably one couched in a symbolic representational system, then frequency and grammar will be dissociable in principle. Bader and Häussler (2010b, c) have shown precisely this: being ungrammatical does entail being low frequency, but being low frequency does not entail being ungrammatical (see also Arppe & Järvikivi, 2007). In effect, rather than not doing something systematically, speakers will systematically not do something.

Perhaps the most crucial property of being a grammatical constraint is that the constraint is *de-coupled from its source*. Imagine a constraint that emerges due to processing complexity. Once it has been grammaticalized, its nature is settled in the grammar, and violation of the constraint should give rise to degraded acceptability whenever the constraint is violated, not just in the complex cases that may originally have motivated the constraint. This is the basis for our argument that the degraded acceptability of simple six-word English multiple questions provides strong evidence for a grammatical Superiority constraint. Arguably this is also what we see in the relative clause island study.

---

[11] Haider (2000) argues along the same line to account for the observation that identity of Case has the same degrading effect in crossing movement for pairs of objects:

(i) *Wen$_i$ hat er denn wen gebeten [davon t_i abzuhalten]?

Further, despite differences in the acceptability of Superiority violations across English and German, there is no difference in the island violation penalty. What interests us here most is the logic: once a prohibition has become part of the grammar, it should have a life of its own and the acceptability of sentences should not necessarily reflect whatever processing or other difficulty might have motivated the constraint initially. The interplay between the decrement in acceptability due to ungrammaticality might be larger than any decrement due to processing difficulty, or with large processing costs, perhaps both decrements will be evident. Assuming that actual linguistic data reflect both the grammar and the processor, in complex or difficult-to-process sentences, one may see effects of both grammar and processing. It is for this reason that sentences which are simple to process may provide the purest reflection of grammar. The badness of complex sentences might be attributed to either source.

The present work is an attempt to address empirically one vexing issue concerning the proper characterization of Superiority in English and German. Every study reported here, and in our earlier work, consistently indicates that the grammar of English contains a constraint prohibiting extraction of the lower of two *wh*-s. In German, it is clear that the status of Superiority is different. The violation costs are smaller than in English (Experiment 1), and they interact with animacy in a different fashion than in English (like animacy of the *wh*-s matters more than in English). Although there is much that remains to be done in choosing the correct grammatical characterization of the English facts, and the appropriate non-grammatical account of the German facts (at least those involving distinct-animacy examples), the present work does help to

---

(ii) Wen$_i$ hat er denn wem versprochen [davon t_i abzuhalten]?

determine the kind of account needed for each language. Further, the present results argue against a parsing account that attributes the weaker and more variable effects of Superiority found in German to differences in dependency length (Experiments 1 and 4). The results of the present studies also show completely comparable island violations in English and German in the same studies showing distinct Superiority effects in the two languages, indicating the undesirability of maintaining that both island constraints and Superiority are due to processing difficulty alone. In conclusion, it means something to say a constraint is 'in the grammar.' This isn't just loose talk. The claim has empirical ramifications.

References

Arnon, I., N. Snider, P. Hofmeister, T. F. Jaeger, & I. Sag. 2006 . Cross-linguistic variation in a processing account: Multiple *wh*-questions. In *Proceedings of Berkeley Linguistics Society, Volume 32*. Berkeley Linguistics Society.

Arppe, A. & J. Järvikivi. 2007. Every method counts: Combining corpus-based and experimental evidence in the study of synonymy. *Corpus Linguistics and Linguistic Theory* 3: 131–159.

Bader, M., & J. Bayer. 2006. *Case and linking in language comprehension: Evidence from German*. Heidelberg: Springer.

Bader, M. & J. Häussler. 2010a. Word order in German: A corpus study. *Lingua 120*: 717–762.

Bader, M. & J. Häussler. 2010b. Toward a model of grammaticality judgments. *Journal of Linguistics* 46: 273–330.

Bader, M. & J. Häussler. 2010c. The primacy of grammaticality. Manuscript, University of Konstanz.

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

Bornkessel, I. &M. Schlesewsky. 2006. The Extended Argument Dependency Model: A neurocognitive approach to sentence comprehension across languages. *Psychological Review 113*: 787-821.

Chomsky, N. 1973. Conditions on transformations. In *A festschrift for Morris Halle*, ed. S. Anderson and P. Kiparsky, 232-286. New York: Holt, Rinehart and Winston.

Chomsky, N. 1981. *Lectures on Government and Binding*. Dordrecht: Foris

Chomsky, N. 2008. On phases. In *Foundational issues in linguistic theory. Essays in honor of Jean-Roger Vergnaud*, eds. R. Freidin, C. P. Otero, and M. L. Zubizarreta, 133-166. Cambridge, MA: MIT-Press.

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

Fanselow, G. 2001. Features, θ-roles, and free constituent order. *Linguistic Inquiry* 32: 405-437.

Fanselow, G. 2010. Semantic type effects on crossing movement in German. *In Language and Logos. A festschrift for Peter Staudacher*. ed. T. Hanneforth and G. Fanselow, 45 - 62. Berlin: Akademie-Verlag.

Fanselow, G. & Féry, C. 2008. Missing superiority effects. Long movement in German (and other languages). In *Elements of Germanic and Slavic grammars. A comparative view*. ed. J. Witkos & G. Fanselow, 67-87. Frankfurt: Lang.

Fanselow, G., M. Schlesewsky, R. Vogel & T. Weskott. 2011. Animacy effects on crossing *wh*-movement in German. *Linguistics* 49: 657–683.

Featherston, S. 2005. Universals and grammaticality: *Wh*-constraints in German and English. *Linguistics* 43: 667-711.

Fedorenko, E. & E. Gibson. 2008. Syntactic parallelism as an account of superiority effects: Empirical investigations in English and Russian. Manuscript. MIT, Cambridge, MA.

Fodor, J. D. 1978. Parsing strategies and constraints on transformations. *Linguistic Inquiry* 9: 427-473.

Gennari, S., & M.C. MacDonald. 2009. Linking production and comprehension: The case of relative clauses. *Cognition* 111: 1-23.

Gibson, E. 2000. The dependency locality theory: a distance-based theory of linguistic complexity. In *Image, language, brain*, ed. Y. Miyashita, A. P. Marantz and W. O'Neil, 95-126. Cambridge, MA: MIT Press.

Grant, M., Clifton Jr., C., & L. Frazier. 2012. The role of non-actuality implicatures in processing elided constituents. *Journal of Memory and Language* 66: 326-343.

Grewendorf, G. 1989. *Ergativity in German*. Dordrecht: Foris.

Haider, H. 1993. *Deutsche Syntax, generativ - Vorstudien zur Theorie einer projektiven Grammatik*. Tübingen: Narr.

Haider, H. 2000. Towards a superior account of superiority. In *Wh-scope marking*. ed. U. Lutz, G. Müller and A. von Stechow, 231-248. Amsterdam: Benjamins.

Hawkins, J. 1994. *A performance theory of order and constituency*. Cambridge: Cambridge University Press.

Hawkins, J. 2004. *Efficiency and complexity in grammars*. Oxford: Oxford University Press.

Hoberg, U. 1981. *Die Wortstellung in der geschriebenen deutschen Gegenwartssprache*. München: Hueber.

Hofmeister, P. 2007. Retrievability and gradience in filler-gap dependencies. In *Proceedings of the 43rd Regional Meeting of the Chicago Linguistics Society*.

Hofmeister, P., T. F.Jaeger, I. Arnon, I. A. Sag & N. Snider (2013). The source ambiguity problem: Distinguishing the effects of grammar and processing on acceptability judgments. *Language and Cognitive Processes* 28: 48-87.

Hofmeister, P., T. F. Jaeger, I. A. Sag, I. Arnon & N. Snider. 2007. Locality and accessibility in *wh*-questions. In *Roots: Linguistics in search of its evidential base*. ed. S. Featherston & W. Sternefeld, 185-206. Berlin: Mouton de Gruyter.

Hofmeister, P. & Sag, I. 2010. Cognitive constraints and island effects. *Language* 86: 366-415.

Hofmeister, P., L. Staum-Casasanto & I. A. Sag. 2010. Understanding acceptability judgments: distinguishing effects of grammar and processing. Poster presented at the 23rd CUNY conference on Human Sentence Processing. March 18-20, 2010. New York University, New York, NY.

Jurka, J. 2010. The importance of being a complement. CED-effects revisited. Ph.d. Dissertation. University of Maryland, College Park.

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

Kempen, G.& K. Harbusch. 2005. The relationship between grammaticality ratings and corpus frequencies: a case study into word-order variability in the midfield of German clauses. In *Linguistic Evidence*. Ed. M. Reis, M. & S. Kepser, 329-349. Berlin: Mouton de Gruyter.

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

Kuno, S. & J. Robinson. 1972. Multiple *wh*-questions. *Linguistic Inquiry* 3: 463-487.

Lawrence, M.A. 2011. ez: Easy analysis and visualization of factorial experiments. R package version.3.0-0.

MacWhinney, B., E. Bates & R. Kliegl. 1984. Cue validity and sentence interpretation in English, German, and Italian. *Journal of Verbal Learning and Verbal Behavior* 23: 127–150.

Mak, W. M., W. Vonk & H. Schriefers. 2002. The influence of animacy on relative clause processing. *Journal of Memory and Language* 47: 50–68.

Mak, W. M., W. Vonk & H. Schriefers. 2006. Animacy in processing relative clause: The hikers that rocks crush. *Journal of Memory and Language* 54: 466–490.

Meng, M. & M. Bader. 2000. Mode of disambiguation and garden path strength: An investigation of subject-object ambiguities in German. *Language and Speech* 43: 43-74.

Meyer, R. 2004. *Syntax der Ergänzungsfrage. Empirische Untersuchungen am Russischen, Polnischen und Tschechischen*. München: Sagner.

Pesetsky, D. 1982. *Paths and categories*. Doctoral dissertation, MIT. Cambridge, MA.

Pesetsky, D. 1987. *Wh*-in-situ: Movement and unselective binding. In *The representation of (in) definiteness*, ed. E. Reuland and A. ter Meulen, 98-129. Cambridge, MA: MIT Press.

Pesetsky, D. 2000. *Phrasal movement and its kin*. Cambridge, Mass: MIT Press.

Phillips, C. 2013. Some arguments and nonarguments for reductionist accounts of syntactic phenomena. *Language and Cognitive Processes* 28: 156-187.

R Development Core Team. 2009. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing.

Reinhart, T. 1995. *Interface strategies*. OTS Working papers in Linguistics.

Ross, J. R. 1967. *Constraints on variables in syntax*. Doctoral dissertation, Cambridge, Mass.

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

Sorace, A. & F. Keller. 2005. Gradience in linguistic data. *Lingua* 115: 1497-1524.

Traxler, M. J., R. Morris & R. Seely. 2002. Processing subject and object relative clauses: evidence from eye movements. *Journal of Memory and Language* 47: 69–90.

Traxler, M. J. & M. J. Pickering. 1996. Case-marking in the parsing of complement sentences: Evidence from eye movements. *Quarterly Journal of Experimental Psychology* 49A: 991-1004.

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

Wiltschko, M. 1997. D-Linking, scrambling, and superiority in German. *Groninger Arbeiten zur Germanistischen Linguistik* 41: 107-142.

Appendix A. Materials Experiment 1

a. English

The four versions of each item can be derived as exemplified for the first item.

1. a. The janitor kept tabs on who smuggled who into this dorm.
   b. The janitor kept tabs on who smuggled what into this dorm.
   c. The janitor kept tabs on who who smuggled into this dorm.
   d. The janitor kept tabs on what who smuggled into this dorm.
2. We found out who hid {who|what} after the riot.
3. Nobody could remember who cursed {who|what} after the local election.
4. The police chief wouldn't reveal who discovered {who|what} after the blackmail attempt.
5. Thomas didn't say who admired {who|what} at the computer fair.
6. The inspector wouldn't let on who photographed {who|what} during the summit.
7. The instructor wouldn't say who mentioned {who|what} in the discussion group.
8. The authorities wouldn't reveal who ridiculed {who|what} after the soccer matc
9. The neighbors laid out who described {who|what} during the investigation.
10. Florian was reminded who mentioned {who|what} at the press release.

11. The professor was surprised by who praised {who|what} at the student workshop
12. The crimes reporter knew who rescued {who|what} last year.
13. Francois told us who complimented {who|what} after the soccer match.
14. The scientist couldn't remember who described {who|what} in the cafeteria.
15. The publicist mentioned who videotaped {who|what} at the award ceremony.
16. The diplomat couldn't recall who overlooked {who|what} at the important international ceremony.
17. The employees knew exactly who criticized {who|what} at the office.
18. The helicopter pilots were uncertain who hid {who|what} inside enemy territory.
19. We forgot who mentioned {who|what} during the soccer tournament.
20. The witness couldn't remember who cursed {who|what} at the scene of the accident.

b. German

The four versions of each item can be derived as shown for the first item below. Translations correspond to the English sentences above. For reason of space we droped the matrix clause.

1. Der Hausmeister führte Listen darüber, wer {wen|was} ins Wohnheim schmuggelte.
   The janitor kept lists on who who what into-the dorm smuggled
   a. Der Hausmeister führte Listen darüber, wer wen ins Wohnheim schmuggelte.
      The janitor kept lists on who who into-the dorm smuggled
   b. Der Hausmeister führte Listen darüber, wer was ins Wohnheim schmuggelte.
      The janitor kept lists on who what into-the dorm smuggled

c. Der Hausmeister führte Listen darüber, wen wer ins Wohnheim schmuggelte.

The janitor kept lists on who who into-the dorm smuggled

d. Der Hausmeister führte Listen darüber, was wer ins Wohnheim schmuggelte.

The janitor kept lists on what who into-the dorm smuggled

2. …, wer {wen|was} nach dem Aufstand versteckt hielt.

who who what after the riot hidden kept

3. …, wer {wen|was} nach der Kommunalwahl verflucht hat.

who who what after the local-election cursed has

4. …, wer {wen|was} nach dem Erpressungsversuch entdeckt hat.

who who what after the blackmail-attempt discovered has

5. …, wer {wen|was} auf der Computermesse bewundert hat.

who who what at the computer-fair admired has

6. …, wer {wen|was} während des Gipfeltreffens fotografiert hat.

who who what during the summit photographed has

7. …, wer {wen|was} in der Diskussionsgruppe erwähnt hat.

who who what in the discussion-group mentioned has

8. …, wer {wen|was} nach dem Fußballspiel verspottet hat.

who who what after the soccer-match ridiculed has

9. …, wer {wen|was} während der Untersuchung beschrieben hat.

who who what during the investigation described has

10. …, wer {wen|was} bei der Pressekonferenz erwähnt hat.

who who what at the press-release mentioned has

11. …, wer {wen|was} bei der studentischen Tagung gelobt hat.

who who what at the student workshop praised has

12. …, wer {wen|was} letztes Jahr gerettet hat.

who who what last year rescued has

13. …, wer {wen|was} nach dem Fußballspiel gelobt hat.

who who what after the soccer-match complimented has

14. …, wer {wen|was} in der Cafeteria beschrieben hat.

who who what in the cafeteria described has

15. …, wer {wen|was} während der Preisvergabe gefilmt hat.

who who what during the award-ceremony videotaped has

16. …, wer {wen|was}bei dem wichtigen internationalen Empfang übersehen hatte.

who who what at the important international reception overlooked had

17. …, wer {wen|was} im Büro kritisierte.

who who what in-the office criticized

18. …, wer {wen|was} im Feindgebiet versteckt hat.

who who what inside enemy-territory hidden has

19. …, wer {wen|was} während des Fußballturniers erwähnte.

who who what during the soccer-tournament mentioned

20. …, wer {wen|was} am Unfallort beschimpft hat.

who who what on-the scene-of-the-accident cursed has

Appendix B. Materials Experiment 2

a. English

Each item is shown in the complement-clause condition; the matrix verb for the relative-clause conditions is given in brackets. The four versions of each item can be derived as exemplified for the first item.

1 a. Jason persuaded the girl to write a large check.

b. Jason met the girl who wrote a large check.

c. What did Jason persuade the girl to write?

d. What did Jason meet the girl who wrote?

2 Karl urged the politician to support long paternity leaves. [interviewed]

3 Heinrich forced the butcher to sell old meat. [criticized]

4 John expected the mayor to oppose rapid development. [praised]

5 Leah suspected the visitor to have stolen the gold statue. [watched]

6 Lisa watched the pedestrian push a baby carriage. [bumped]

7 Carol urged the gardener to fertilize the tea roses. [noticed]

8 Caroline convinced the transfer student to study Calculus. [helped]

9 Hans asked the neighbors to buy a garbage container. [liked]

10 William asked the volunteer to clean the big pots. [telephoned]

11 Kyle invited the new teacher to coach basketball. [disliked]

12 Thomas accused the student of copying the term paper. [accused]

13 Angela coerced the bureaucrat into distributing reimbursement forms. [antagonized]

14 Brenda helped the farmer to raise horses. [helped]

15 Tina believed the boy to have thrown spitballs. [pinched]

16 Paula advised the fisherman to use a large net. [kissed]

b. German

The four versions of each item can be derived as exemplified for the first item below. Translations correspond roughly to the English sentences above.[12]

1 a. Jakob überredete das Mädchen einen großen Scheck auszustellen.

J. persuaded the girl a large check to- write-out

b. Jakob traf das Mädchen, das einen großen Scheck ausgestellt hatte.

J. met the girl who a large check out-written had

c. Was überredete Jakob das Mädchen auszustellen?

What persuaded J. the girl to-write-out

d. Was traf Jakob das Mädchen, das ausstellte?

What met J. the girl who out-wrote

2 Karl drängte den Politiker längere Elternzeiten zu unterstützen. [interviewte]

K. urged the politician longer paternity-leaves to support interviewed

[12] An exception is the relative-clause version of (6) which contains a matrix verb that deviate from the English counterpart. This deviation is due to the fact that the German counterpart for English *bump* is a particle verb and our decision to avoid particle verbs in German.

3 Heinrich zwang den Metzger verdorbenes Fleisch zu verkaufen. [kritisierte]

H. forced the butcher tainted meat to sell critized

4 John ersuchte den Bürgermeister die übereilten Planungen abzulehnen. [lobte]

J. asked the mayor the overhastly plannings to-reject praised

5 Leah verdächtigte den Besucher die goldene Statue gestohlen zu haben. [beobachtete]

L. suspected the visitor the golden statue stolen to have watched

6 Lisa sah den Fußgänger einen Kinderwagen schieben. [warnte]

L. saw the pedestrian a baby-carriage push warned

7 Carola drängte den Gärtner die Teerosen zu düngen. [bemerkte]

C. urged the gardener the tea-roses to fertilize noticed

8 Caroline überredete den Austauschstudenten den Logikkurs zu belegen. [unterstützte]

C. convinced the transfer-student the logics-course to attend helped

9 Hans bat die Nachbarn einen neuen Müllcontainer zu kaufen. [mochte]

H. asked the neighbors a new garbage-container to buy liked

10 Wilhelm bat den Praktikanten die großen Töpfe zu reinigen. [rief]

W. asked the volunteer the big pots to clean called

11 Kilian verpflichtete den neuen Lehrer Basketball zu unterrichten. [verachtete]

K. obliged the new teacher basketball to teach despised

12 Thomas beschuldigte den Studenten die Hausarbeit abgeschrieben zu haben. [verwarnte]

T. accused the student the term-paper copied to have cautioned

13 Angela nötigte den Bürokraten die Entschädigungsformulare zu verteilen. [ärgerte]

A. coerced the bureaucrat the reimbursement-forms to distribute antagonized

14 Britta ermunterte den Bauern Pferde zu züchten. [ermunterte]

B. encouraged the farmer horses to breed encouraged

15 Tina beschuldigte den Jungen Papierkügelchen geworfen zu haben. [ohrfeigte]

T. accused the boy paper-balls thrown to have ear-boxed

16 Paula lehrte den Fischer ein großes Netz zu benutzen. [küsste]

P. taught the fisherman a large net to use kissed

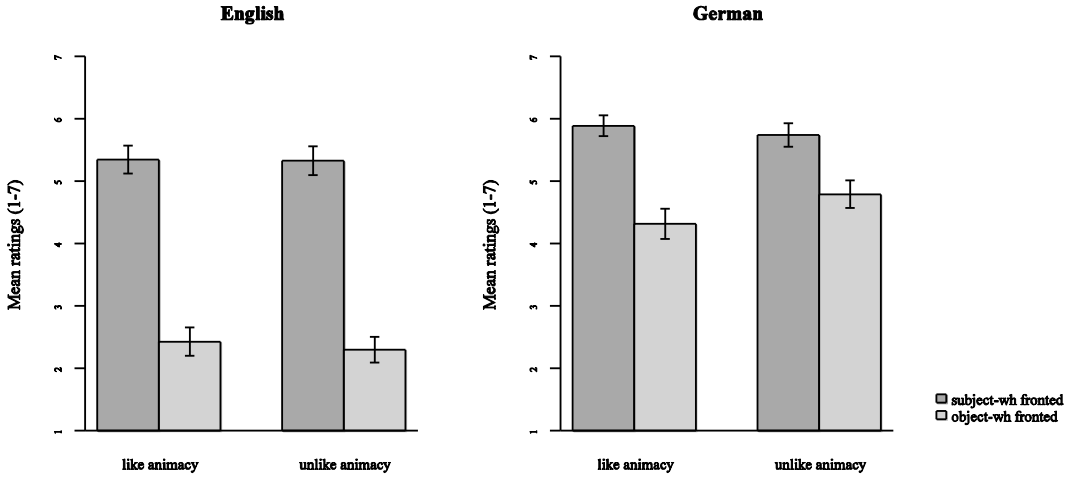

Figure 1. Mean ratings in Experiment 1 as a function of order among the two *wh*-s and their animacy. Vertical bars indicate 95% confidence intervals.

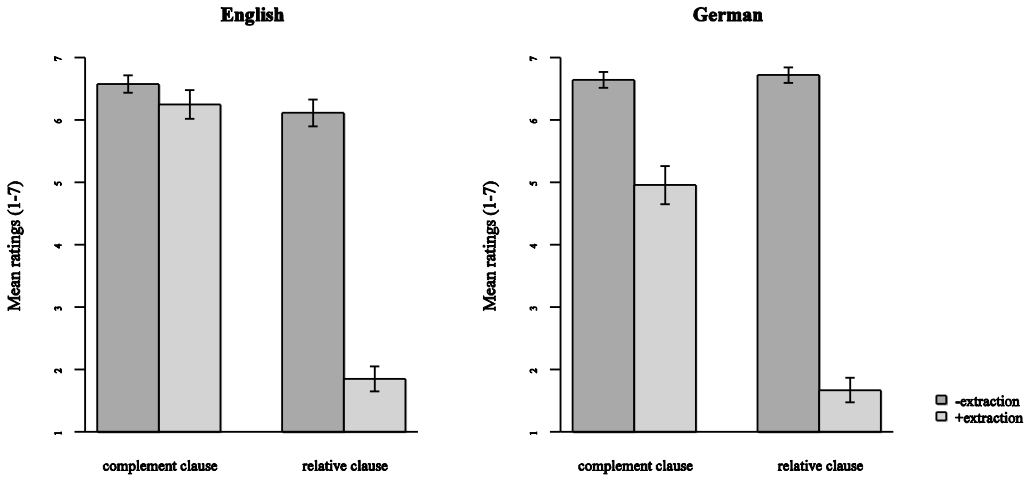

Figure 2. Mean ratings in Experiment 2 as a function of clause type and extraction. Vertical bars indicate 95% confidence intervals.

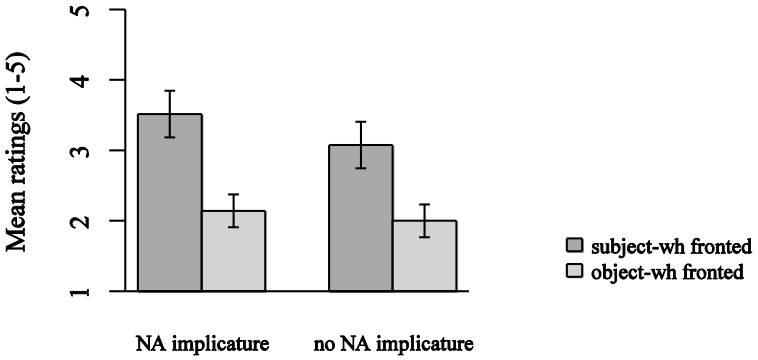

Figure 3. Mean ratings in Experiment 3 as a function of context (+/- containing an NAI-inducing element) and extraction. Vertical bars indicate 95% confidence intervals.

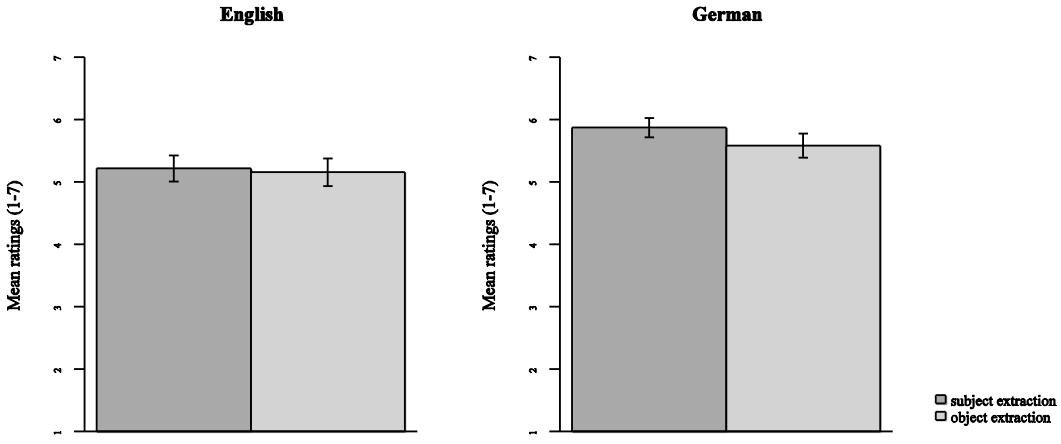

Figure 4. Mean ratings in Experiment 4 as a function of Extraction. Vertical error bars indicate 95% confidence intervals.