

# Superiority Effects\*

Yehuda N. Falk

Department of Linguistics

The Hebrew University of Jerusalem

msyfalk@mscc.huji.ac.il

<http://pluto.mscc.huji.ac.il/~msyfalk/>

## 1. The Problem

*Wh* questions with multiple *wh* elements pose a special problem within the overall analysis of *wh*-type constructions. In a language like English, which only allows one *wh* element to be in the fronted discourse-prominent position, any additional *wh* element occupies an “in situ” position; i.e. the position of the clause-internal grammatical function.<sup>1</sup> As noted at least since Kuno and Robinson (1972), the choice of which *wh* element is fronted is not free. They cite examples such as the following:

- (1) a. Who did what?  
b. Who went where?  
c. What happened to whom?  
d. What did you give to whom?
- (2) a. \*What did who do?  
b. \*Where did who go?  
c. \*To whom did what happen?  
d. \*To whom did you give what? / \*Who did you give what to?

A similar effect obtains in (a subset of) languages which front multiple *wh* elements, as seen in the following examples from Bulgarian (Bošković 2002: 354).

- (3) a. Koj kogo običa ?  
who whom loves  
'Who loves whom?'  
b. \*Kogo koj običa ?  
whom who loves

Expressing the constraint in movement terms, Kuno and Robinson state that a *wh* element cannot cross over another *wh* element. Stated this way, the constraint seems rather arbitrary. It becomes even more arbitrary when compared with cases such as the following:

- (4) a. Where did you buy what? / What did you buy where?  
b. When did you buy what? / What did you buy when?

---

\*Draft, February 2012. This research is supported by the Israel Science Foundation, grant #207/07.

<sup>1</sup>I will use the term *in situ* in this paper to refer to *wh* elements in the internal-function position because of its familiarity. However, I consider the term inappropriate: each of the two possible positions of the *wh* element is in place for one of its functions.

Unlike in the previous cases, here either *wh* can be fronted. Kuno and Robinson propose an arbitrary exemption for expressions of time and place.

Since Chomsky (1973), this effect has been referred to as Superiority. Superiority effects have been extensively discussed in the Principles and Parameters literature as a syntactic locality condition (ECP, Minimal Link, Attract Closest). Outside the P&P literature, Superiority effects have received a good deal less attention, and the status of Superiority as a syntactic condition has been questioned.

The thesis of the present study is that Superiority is partially, but not completely, syntactic. Involving, *ex hypothesi*, more than one component of grammar, it is most naturally accounted for in a multidimensional theory of language, such as Lexical-Functional Grammar. In §2, I will discuss reasons for questioning the syntactic status of Superiority. I will argue that extrasyntactic factors alone cannot account for the effect, but that syntax alone cannot account for it either. In §3, I will propose a multidimensional LFG account of Superiority, one in which *c*-structure, *f*-structure, and *i*-structure all play a role. Finally, in §4, I will consider the question of why the Superiority effect exists.

## 2. Superiority: Syntax or Not

### 2.1. Pragmatics

Since Superiority effects are evident in *wh* questions, a reasonable hypothesis is that what is involved is pragmatics rather than syntax. This is, in part, the position of Bolinger (1978), who argues that given the correct context, any order is possible. He cites the following example.

- (5) a. Who broke what?  
 b. \*What did who break?  
 c. I know that among all the disasters in that kitchen, Jane scorched the beans and Lydia put salt in the ice tea; but *whát* did *whó* *bréak*? I know somebody broke something, so stop evading my question.

(5a,b) illustrate the usual Superiority effect. In (5b) the fronted *what* has (in transformational terms) moved over *who*, and the result is ungrammatical. However, in (5c), the same thing has happened, and yet the result is grammatical. Although he does not state it explicitly, the *wh* elements here do not represent completely novel elements in the discourse. A context has been created in which talking about somebody breaking something is natural. It is this connection to the context, apparently, which makes this grammatical.

The idea that lack of complete novelty makes Superiority violations grammatical has become fairly well established in the literature on Superiority. It has been discussed under various names—for example, Arnon, Snider, Hofmeister, Jaeger, and Sag (to appear) and Hofmeister, Jaeger, Sag, Arnon, and Snider (2006) refer to elements that are connected to previous discourse as “accessible.” However, the most common term used in the literature, which I will use here, is *discourse linking*, or *D-linking* (Pesetsky 1987). D-linking has been discussed in the literature primarily in the context of *which* phrases, which appear to be immune to Superiority effects,<sup>2</sup> as in the following examples from Pesetsky (2000: 16).

---

<sup>2</sup>In multiple *wh* fronting languages, this manifests itself a little differently: D-linked *wh* elements need not front (Bošković 2002).

- (6) a. Which person bought which book?  
 b. Which book did which person buy?
- (7) a. Which person did John talk to about which topic?  
 b. Which topic did John talk to which person about?

Pesetsky (2000: 16) characterizes *wh* questions with D-linked elements as ones in which “a *wh*-question asks for answers in which the individuals that replace the *wh*-phrases are drawn from a set that is presumed to be salient to both speaker and hearer”. Pesetsky (1987: 109) also provides examples without *which*, similar to the Bolinger example quoted in (5c).

- (8) a. I know what just about everybody was asked to do, but *what* did *who* (actually) do?  
 b. I know that we need to install transistor A, transistor B, and transistor C, and I know that these three holes are for transistors, but I’ll be damned if I can figure out from the instructions *where what* goes!

While Pesetsky (1987, 2000) attempts to connect the different behavior of D-linked *wh* elements to syntax (admitting (2000: 41) that it is not clear why they should differ syntactically), it is clear that D-linking is not a syntactic notion, but rather a semantic or pragmatic one. I follow Ginzburg and Sag (2000: 248f) in treating D-linking as pragmatic. They argue against the claim by Pesetsky (1987) that D-linked phrases are semantically different (not quantifiers, unlike standard *wh*’s); instead, they take the position that the difference is in the presuppositions, and thus pragmatic in nature.

A promising approach to the pragmatic difference between D-linked and non-D-linked *wh* elements is that of Bošković (2002: 360), who observes that D-linked elements, not being completely novel, are not truly focal. I will adopt this approach here. This represents a departure from the usual LFG view, which treats all question elements as foci. I hypothesize that among the functions at information structure (i-structure), in addition to the standard LFG FOCUS, TOPIC, BACKGROUND.INFO, and COMPLETIVE.INFO (Butt and King 2000), there is also a function Q, which is the function of question elements. Ordinary *wh* elements bear both of the discourse functions Q and FOCUS, while D-linked *wh* elements are Q, but not FOCUS.<sup>3</sup> The i-structure of (1a), with no D-linking, is (9a), while the i-structure of (8a), with D-linking, is (9b).

- (9) a. 
$$\left[ \begin{array}{l} \text{Q} \\ \text{FOCUS} \\ \text{BKGRD.INFO} \end{array} \right] \left\{ \begin{array}{l} \{ [who] \} \\ [what] \\ [did] \end{array} \right.$$
- b. 
$$\left[ \begin{array}{l} \text{Q} \\ \text{BKGRD.INFO} \end{array} \right] \left\{ \begin{array}{l} \{ [who] \} \\ [what] \\ [did] \end{array} \right.$$

What emerges is the generalization (10).

---

<sup>3</sup>I am leaving a couple of issues open here. The first is what discourse function a D-linked question phrase might have in addition to Q. I suspect that it is BACKGROUND.INFO, but I am not certain. The second issue is the consequences to the feature system that has been proposed for i-structure functions, which generally consists of the two binary features [ $\pm$ New] and [ $\pm$ Prominent]. Q is presumably [+Prominent], and apparently newness is undefined for it. If there are more than four discourse functions at i-structure, a richer feature system is required.

- (10) Superiority effects are manifested when two *wh* question phrases bear both the function Q and the function FOCUS at information structure.

This having been said, it is clear that identifying relevant discourse functions is only a first step. More specifically, it does not follow that the constraint responsible for Superiority effects is not (at least partially) syntactic. Since some syntactic factor is clearly involved (apparently, at this stage, linear order or grammatical functions), what is involved is a constraint referring both to the pragmatics and syntax. This point is important. While I agree with the statement by Mycock (2006: 128) that “the time has come to return to Superiority with an open mind and consider what part non-syntactic factors may play in so-called Superiority effects,” I do not agree with her broader conclusion that the effect is not syntactic.

## 2.2. Prosody

As can be seen in examples (5c) and (8) above, prosodic properties are related to the pragmatics of Superiority: in grammatical sentences that violate Superiority, at least one of the *wh* phrases is prosodically prominent. As a result, it has been suggested by several researchers (such as Bolinger 1978, Mycock 2006, Ginzburg and Sag 2000, Zubizarreta 1998) that Superiority effects are a result of improper prosody. For example, Mycock (2006: 270f) proposes that the prosody of *wh* questions reflects the entire sentence being focused: the final element receiving prosodic prominence. In addition, in-situ *wh* phrases are prosodically prominent. With this in mind, consider the sentences in (11).

- (11) a. Who ate what?  
b. \*What did who break?

In (11a), the in-situ *wh* is the final element in the sentence. (11a) thus meets all relevant prosodic constraints, and is well formed. In (11b), on the other hand, the two constraints would grant prominence to two different elements. The sentence is thus ill-formed.

Attractive though it may seem to reduce Superiority effects to prosody, it is not successful. In the first place, it is not clear how it would extend to Superiority effects in multiple-*wh*-fronting languages like Bulgarian ((3) above). Second, even in English it is possible for a second in-situ *wh* to be non-final (Mycock 2006: 267 fn12):

- (12) Who put what on the shelf?

Under the prosodic account, this should be ill-formed. The fact that it is not shows that (11b) is not ill-formed because of a clash of prosodic prominence constraints.

The most important reason to reject a prosodic account of Superiority comes from contrasting (12) with (13).

- (13) \*What did who put on the shelf?

Pronounced with relatively neutral intonation, these two sentences contrast with each other. (13) needs heavy stress on *who*, while *what* in (12) takes a much lighter stress associated with all in-situ *wh* elements, as discussed below. This prosodic contrast is clearly a consequence of Superiority, not its cause.

The prosodic prominence of the in situ *wh* element in apparent Superiority violations is not an independent property, but rather a consequence of D-linking. Ginzburg and Sag (2000:

255ff) includes a study of in situ *wh* in English. While it is generally assumed that only echo questions can involve in situ *wh*, they show that the distribution of in situ *wh* is broader. First of all, in addition to echo uses, such as (14), there are also what they refer to as reference questions, as in (15).<sup>4</sup>

- (14) A: I hear Bustamante y Bacigalupo plays the violin badly.  
 B: You hear WHO plays the violin badly?
- (15) A: I think they're mad at Bustamante y Bacigalupo.  
 B: You think WHO's mad at Bustamante y Bacigalupo?

In these two types of questions, collectively “reprise” questions, the in situ *wh* is heavily stressed.<sup>5</sup> As shown by the term “reprise”, these *wh* elements are connected to the previous discourse, and thus D-linked. The stress in the reference question is the same as in violations of Superiority, showing that the relevant intonational contour is a consequence of noninherent<sup>6</sup> D-linking.

Ginzburg and Sag also discuss non-reprise uses of *wh* in situ. Their examples include the following.

- (16) a. Michael Krasny (addressing a guest—WHO HAS NOT SAID ANYTHING YET—about the interim chief of the US Attorney's office): This is a position that is HOW IMPORTANT in your judgment, Rory?
- b. I've been working here for 14 years. You've been here for HOW LONG? A month?

They do not discuss the intonation of these cases. The examples are presented with small capitals, presumably again to indicate stress. However, there is a clear difference between the prosody in the reprise and non-reprise cases; in the non-reprise examples (where the *wh* is not D-linked) the stress is much weaker.<sup>7</sup> This weaker stress is what one finds in the neutral intonation of questions that conform to the requirements of Superiority.

I conclude then that prosody is not the cause of Superiority effects. Non-inherently D-linked elements are marked prosodically, and this is the prosody that appears to be involved in Superiority.

### 2.3. Processing

A novel proposal concerning Superiority effects is that they are side effects of processing. This proposal is defended in such studies as Arnon, Snider, Hofmeister, Jaeger, and Sag (to appear) and Hofmeister, Jaeger, Sag, Arnon, and Snider (2006). They argue that the following generalizations about processing *wh*-type constructions hold:

---

<sup>4</sup>These are based on Ginzburg and Sag's examples. Their examples consist just of the embedded clauses in my examples, with *who* clause-initial, making them not obviously in-situ.

<sup>5</sup>As Ginzburg and Sag point out, the intonation differs in these two cases: high tone in the echo question and low tone in the ref question.

<sup>6</sup>Which NPs, which are inherently D-linked, are not stressed.

<sup>7</sup>This intonational prominence corresponds to what Mycock (2006) refers to as prosodic focusing.

- (17) (Arnon, Snider, Hofmeister, Jaeger, and Sag to appear)
- 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.

Superiority violations are suboptimal on these parameters. To put it slightly differently, all orders and arrangements of *wh* elements are grammatical, but some are unacceptable because they are hard to process.

Like the prosodic analysis, the processing analysis does not hold up under scrutiny. First of all, as in the case of the prosody, it is reasonable to ask whether the unacceptability is a result of the processing difficulties, or the processing difficulties are a consequence of ungrammaticality. Simply demonstrating that something is difficult to process does not establish that the processing is the cause. It is plausible that the parser is designed to disfavor ungrammatical parses.

More to the point, there are several reasons to question the processing analysis: in particular the claim that distance increases the difficulty of the parse. In the context of Superiority, this is alleged to lead to a preference for (18a) over (18b), and (18c) over (18d), where the intermediate material (which constitutes the distance) is underlined.

- (18) a. What did you give  $\emptyset$  to whom?  
 b. \*To whom did you give what  $\emptyset$ ?  
 c. Who  $\emptyset$  did what?  
 d. \*What did who do  $\emptyset$ ?

The differences in distance are small here. Notably, the distance in the grammatical (18a) is (apparently, since distance is not actually defined) the same as the distance in the ungrammatical (18d). It is hard to imagine a processing model in which the former is easy enough to parse to be acceptable while the latter is not. This casts doubt on the processing approach. In addition, it is difficult to see how this would extend to Bulgarian: for the examples in (3), each version results in a distance of zero for one of the *wh* elements, and a distance of one for the other. Distance between gap and filler would not distinguish between the grammatical and ungrammatical varieties.

The concept of distance to which these authors appeal is potentially problematic from a different perspective. In order to speak of the distance between the filler and gap, the gap must be a structural element. At least one of the authors of these papers, Ivan Sag, is on record as opposing the idea of actual structural gaps (or “traces”). As long ago as Sag and Fodor (1994), Sag argued that there is no structural gap, and that what is bound by the filler is a position in the verb’s argument structure. If so, the relevant distance is not to the positions marked  $\emptyset$  in (18), but to the verb heading the clause. Since in both pairs, both elements are arguments of the same verb, there would be no difference.<sup>8</sup>

Finally, the claim that what matters is the distance between the filler and the gap conflicts with another claim that has been made in the literature about processing of *wh*-type constructions: the claim that nested dependencies are preferred over crossing dependencies. The most commonly cited example is the contrast between (19a) and (19b):

---

<sup>8</sup>It is unclear how the concept of distance would work in a theory where some, but not all, gaps are structurally present (Falk 2007).

- (19) a. Which violin<sub>i</sub> is this sonata<sub>j</sub> easy to play  $e_j$  on  $e_i$ ?  
 b. \*Which sonata<sub>j</sub> is this violin<sub>i</sub> easy to play  $e_j$  on  $e_i$ ?

I will refer to this as the violin-sonata phenomenon. The apparent descriptive generalization is that the two dependencies may not cross, but have to be nested. Baker (1977: 63) stated the constraint explicitly in terms of processing:<sup>9</sup>

- (20) a. As a sentence is processed from left to right, a prospective filler  $y$  is *more current* than a prospective filler  $x$  if  $y$  occurs to the right of  $x$ .  
 b. A prospective filler is assigned to the first unoccupied gap for which it is the most current of the eligible prospective fillers.

Fodor (1978: 448) proposed the following as an anti-ambiguity parsing constraint:

- (21) *The Nested Dependency Constraint (NDC)*  
 If there are two or more filler-gap dependencies in the same sentence, their scopes may not intersect if either disjoint or nested dependencies are compatible with the well-formedness conditions of the language.

Either way, the violin-sonata phenomenon is taken to be the result of processing considerations. Crucially, this is a different constraint from the distance-based approach, and the two constraints make different predictions. The relevant DPs in the violin-sonata sentences are D-linked, which means that Superiority would not apply to them, but from the perspective of the processing-based approach D-linking (“accessibility”) is only a mitigating factor. In the violin-sonata sentences, each version has the filler and gap closer for one pair and farther for the other.

- (22) a. Which violin<sub>i</sub> is this sonata<sub>j</sub> easy to play  $e_j$  on  $e_i$ ? (7 words)  
 b. Which violin<sub>i</sub> is this sonata<sub>j</sub> easy to play  $e_j$  on  $e_i$ ? (3 words)  
 (23) a. \*Which sonata<sub>j</sub> is this violin<sub>i</sub> easy to play  $e_j$  on  $e_i$ ? (6 words)  
 b. \*Which sonata<sub>j</sub> is this violin<sub>i</sub> easy to play  $e_j$  on  $e_i$ ? (4 words)

The total of the two distances is even the same in the two cases. The processing approach thus predicts counterfactually either that both are acceptable, or that both are unacceptable. I will return to the violin-sonata phenomenon in §3.4.

The basic problem is that both the hypothesis that closeness is the concept governing the parsing of filler-gap constructions and that the parser prefers nested dependencies are plausible. Both appear to be supported by particular constructions. However, they cannot both be correct. The real generalization appears to be that what is allowed by the grammar is easier to parse than what is not allowed by the grammar. We thus return to the need for a grammatical account.

### 3. The Superiority Condition

#### 3.1. Preliminary Characterization

I turn now to the task of stating the Superiority Condition. Most analyses have treated

---

<sup>9</sup>For readability, I have changed the terminology: *filler* for Baker’s *tenant* and *gap* for his *address*.

it as a locality constraint. The classical analysis in Government/Binding theory is that it is a consequence of the Empty Category Principle (Chomsky 1981). However, as noted by Hendrick and Rochemont (1988), the ECP analysis predicts incorrectly that only subjects exhibit Superiority effects.

- (24) a. \*What did John tell who that Bill liked?  
 b. \*To whom did you give what? / \*Who did you give what to?

More recently, in the Minimalist Program, Superiority has been attributed to the Minimal Link Condition or Attract Closest (Chomsky 1995, Pesetsky 2000). The core idea is that the *wh* that moves is the one closest to the landing site. This is transparent for languages with single-*wh* fronting, such as English. In the case of multiple-*wh* fronting languages, it is more complicated, but it has been proposed by Bošković (2002) that only the first *wh* element undergoes the relevant kind of fronting, *wh* fronting, while others undergo focus fronting. In common with the processing-based account discussed above, the assumption under a Minimal Link/Attract Closest approach is that the conditioning factor is the distance between filler and gap positions.

I am going to propose a different approach here, one based on grammatical functions rather than on closeness. Specifically:

- (25) a. **Superiority Condition** (first approximation)  
 For multiple elements in a sentence bearing the functions Q and FOCUS at information structure, the order of the *wh* elements reflects their relative prominence on the Relational Hierarchy.
- b. **Relational Hierarchy**<sup>10</sup>  
 SUBJ > OBJ > OBJ<sub>θ</sub> > OBL<sub>θ</sub>

In other words, Superiority is about aligning hierarchies, specifically word order and the Relational Hierarchy. By stating Superiority in terms of the order of the *wh* elements, it accounts for the realization both in single-*wh* fronting languages and in multiple *wh*-fronting languages. Recall that I am hypothesizing that D-linked *wh* question elements bear the function Q but not FOCUS; the condition as stated correctly excludes sentences with at least one D-linked *wh* element.

This approach to Superiority has several important advantages over one based on the distance between fillers and gaps. In the first place, it follows without additional stipulation that adjuncts do not display Superiority effects, as shown in (4) (repeated below as (24)).

- (26) a. Where did you buy what? / What did you buy where?  
 b. When did you buy what? / What did you buy when?

Since the Relational Hierarchy only ranks argument-expressing grammatical functions,

---

<sup>10</sup>The grammatical function OBJ<sub>θ</sub> is the function of secondary objects (Bresnan and Kanerva 1989).

adjuncts are not ranked.<sup>11</sup>

A second advantage of basing Superiority on grammatical functions rather than linear order is that not all languages that have Superiority effects have a fixed linear order. For example, we have seen that Bulgarian displays Superiority effects. However, word order in Bulgarian is relatively free (Jaeger and Gerassimova 2002), casting in doubt an approach based on proximity of the gap to the filler.

Finally, the two approaches make different predictions about two *wh* elements with grammatical functions that are equally ranked. This happens when there are two arguments with oblique functions. The distance-based approach predicts that these are no different from any other structure: the closer *wh* can front grammatically, while the second one cannot. On the other hand, the grammatical function-based approach predicts that either *wh* should be able to front. The judgments are subtle, but support the grammatical function-based approach.

- (27) a. Who did you talk with *e* about what?  
 b. (?)What did you talk with who about *e*?  
 c. What did you talk about *e* with whom?  
 d. (?)Who did you talk about what with *e*?

There is a slight degradation when the fronted *wh* is associated with the farther gap, but not the ungrammaticality of Superiority violations.<sup>12</sup> This shows that Superiority is not the result of distance, but rather the Relational Hierarchy.

This account of the Superiority Condition is strikingly multidimensional in character, matching up information structure, functional structure, and constituent structure. Such a multidimensional account is natural in a theory such as LFG, in which language is conceived of as involving a multiplicity of representations, each encoding a particular kind of grammatical information, linked through correspondence rules.

---

<sup>11</sup>The situation is a little more complicated. In English, for example, if the adjunct *why* is one of the *wh* elements it has to be fronted, as in the following example from Hendrick and Rochemont (1988: 86).

- (i) a. Mary remembers *why* John bought what.  
 b. \*Mary remembers what John bought *why*.

According to Bošković (1997: 234), in Bulgarian fronted objects have to precede fronted adjuncts for some speakers. He provides the following examples.

- (ii) a. Kogo kak e tselunal Ivan ?  
 whom how is kissed Ivan  
 'Whom did Ivan kiss how?'  
 b. ?\*Kak kogo e tselunal Ivan ?  
 how whom is kissed Ivan  
 'How did Ivan kiss whom?'  
 c. Kogo küde e vidjal čovetküt ?  
 whom where is seen the.man  
 'Who did the man see where?'  
 d. ???Küde kogo e vidjal čovetküt ?  
 where whom is seen the.man  
 'Where did the man see whom?'

However, in footnote 19 he states that for some speakers (iib) is better than the ?\* marking indicates (although still worse than (iia)), and (iic), which is marked as a weaker violation, is perfectly grammatical. This suggests that the effect here is not due to Superiority. This may be true of the *why* case in English as well.

<sup>12</sup>This slight degradation may be due to processing-based distance or crossing effects, but if so it shows that the result is much weaker than the ungrammaticality of Superiority violations.

### 3.2. Refining the Characterization

Thus far I have only considered structures with two *wh* elements. Widening the investigation to structures with three allows a refinement of the statement of the Superiority Condition.

I begin with Bulgarian. In Bulgarian, when three *wh* elements are fronted, only the first one is subject to Superiority. The other two may occur in any order (Bošković 1997: 239).

- (28) a. Kogo kakvo e pital Ivan ?  
 whom what is asked Ivan  
 ‘Whom did Ivan ask what?’  
 b. ?\*Kakvo kogo e pital Ivan ?  
 what whom is asked Ivan  
 c. Koj kogo kakvo e pital ?  
 who whom what is asked  
 ‘Who asked whom what?’  
 d. Koj kakvo kogo e pital ?  
 who what whom is asked

This leads to a refinement of the original statement of the Superiority Condition.

(29) **Superiority Condition** (second approximation)

For multiple elements in a sentence bearing the functions Q and FOCUS at information structure, the left-most *wh* element is the most prominent on the Relational Hierarchy.

The situation in English is less clear. Kayne (1983: 235) notes an amelioration of Superiority when there are three *wh*'s:

- (30) a. \*I'd like to know where who hid it.  
 b. \*I'd like to know what who hid there.  
 (31) a. ?I'd like to know where who hid what.  
 b. ?I'd like to know what who hid where.

Pesetsky (2000: 17) goes further and states that when there are three *wh* elements there is “no detectable Superiority effect”. The judgments are delicate; multiplying *wh* elements makes the sentences more difficult in any case. If these judgments are correct, it could be the case that the Superiority Condition in English is different, and only applies when there are exactly two *wh* elements. However, there may be a complicating factor here: a D-linked interpretation. Impressionistically, the sentences in (29) have D-linked stress on the offending *wh* elements. This prosodic effect is assisted by the fact that one of the *wh* elements is in sentence-final position. Adding material at the end of the sentence appears to make the sentences worse, although perhaps not as bad as normal Superiority violations.

- (32) a. ?\*I'd like to know where who hid what yesterday.  
 b. ?\*I'd like to know what who hid where yesterday.

The presence of three *wh* elements may pragmatically invite an interpretation where one or more are D-linked. If this is the case, elements with three *wh* elements are not true exceptions to the Superiority Condition.

### 3.3. Languages without Superiority Effects

There appear to be languages that do not exhibit Superiority Effects, or only exhibit them in certain contexts. One such language that has been cited extensively in the literature is Spanish (Chomsky 1981: 255):

- (9) a. ¿ Quién compró qué ?  
       who bought what  
       ‘Who bought what?’  
       SUBJ < OBJ
- b. ¿ Qué compró quién ?  
       what bought who  
       OBJ < SUBJ
- c. Juan sabe qué dijo quién.  
       Juan knows what said who  
       OBJ < SUBJ  
       ‘Juan knows who said what.’
- d. Juan sabe quién dijo qué.  
       Juan knows who said what  
       SUBJ < OBJ

Similarly, German appears to lack Superiority effects (Pesetsky 2000: 17):

- (33) a. Ich weiß nicht, wer was gesehen hat.  
       I know not who what seen has  
       ‘I don’t know who has seen what.’  
       SUBJ < OBJ
- b. Ich weiß nicht, was wer gesehen hat.  
       I know not what who seen has  
       ‘I don’t know what who has seen.’  
       OBJ < SUBJ

Among multiple-fronting languages, Russian has no Superiority (Bošković 2002: 354):

- (34) a. Kto kogo ljubit ?  
       who whom loves  
       ‘Who loves whom?’  
       SUBJ < OBJ
- b. Kogo kto ljubit ?  
       whom who loves  
       OBJ < SUBJ

Unlike researchers in the Principles and Parameters tradition, I see no reason not to simply hypothesize that the Superiority Condition is not operative in some languages.<sup>13</sup>

In some languages, Superiority is operative only when the *wh*-path crosses a clause

---

<sup>13</sup>Even in P&P, some difference between languages needs to be postulated. For example, Pesetsky (2000) hypothesizes that English and German differ in what kinds of inaudible complementizers each one has in the lexicon. This hardly seems an improvement over the approach that I am taking; it is certainly no more explanatory.

boundary. Thus while within a clause, as noted above, there are no Superiority effects in German, they emerge across clausal boundaries (Pesetsky 2000: 77).

- (35) a. Wer glaubte, daß der Peter ihr wen vorstellte ?  
 who.NOM believed that the Peter her.DAT whom.ACC introduced  
 ‘Who believed that Peter introduced whom to her?’  
 SUBJ < OBJ
- b. ?\*Wen glaubte wer, daß der Peter ihr vorstellte ?  
 whom.ACC believed who.NOM that the Peter her.DAT introduced  
 OBJ < SUBJ

This is also true in Serbo-Croatian (Bošković 2002: 353f).

- (36) a. Ko koga voli ?  
 who whom loves  
 ‘Who loves whom?’  
 SUBJ < OBJ
- b. Koga ko voli ?  
 whom who loves  
 OBJ < SUBJ
- (35) a. Ko koga kažeš da je istukao ?  
 who whom say that is beaten  
 ‘Who do you say beat whom?’  
 SUBJ < OBJ
- b. \*Koga ko kažeš da je istukao ?  
 whom who say that is beaten  
 OBJ < SUBJ

We can update our statement of the Superiority Condition.

(37) **Superiority Condition** (final statement)

For multiple elements in a sentence bearing the functions *Q* and *FOCUS* at information structure <at least one of them in a different clause from the clause-internal grammatical function>, the left-most *wh* element is the most prominent on the Relational Hierarchy.

I note in passing that in some cases what looks like lack of Superiority effects may result from multiple *wh* elements not bearing the *FOCUS* function, or some other *i*-structure factor. For example, Hungarian, a multiple *wh*-fronting language, lacks Superiority effects, but the first *wh* element is D-linked (Anna Gazdik, personal communication), and thus not *FOCUS*. It thus follows from my analysis that Superiority effects will be absent. Pesetsky (2000) notes a correlation between lack of Superiority effects and the ability to give a single-pair answer to multiple *wh* questions. While the source of the distinction between pair-list and single-pair readings is not clear, *i*-structure status of the *wh* elements is plausibly involved.

### 3.4. On Violins and Sonatas

In §2.3, mention was made of a second constraint on multiple *wh*-type dependencies, the “violin-sonata phenomenon.” The classic example, given above as (19), is repeated here.

- (38) a. Which violin<sub>*i*</sub> is this sonata<sub>*j*</sub> easy to play *e<sub>j</sub>* on *e<sub>i</sub>*?  
 b. \*Which sonata<sub>*j*</sub> is this violin<sub>*i*</sub> easy to play *e<sub>j</sub>* on *e<sub>i</sub>*?

The constraint is usually characterized as requiring nested dependencies rather than crossing dependencies. Here are a few more examples culled from the literature.

- (39) a. Which city<sub>*i*</sub> is Susan<sub>*j*</sub> tough to drive {with *e<sub>j</sub>* to *e<sub>i</sub>* | \*to *e<sub>i</sub>* with *e<sub>j</sub>*}?  
 b. Who<sub>*i*</sub> is Chicago<sub>*j*</sub> tough to drive {to *e<sub>j</sub>* with *e<sub>i</sub>* | \*with *e<sub>i</sub>* to *e<sub>j</sub>*}?  
 (Culicover and Wexler 1977: 27)

- (40) a. What<sub>*i*</sub> are boxes<sub>*j*</sub> easy to store *e<sub>j</sub>* in *e<sub>i</sub>*?  
 b. \*What<sub>*i*</sub> are boxes<sub>*j*</sub> easy to store *e<sub>i</sub>* in *e<sub>j</sub>*?  
 (Fodor 1978: 448)

- (41) a. These girls<sub>*i*</sub>, these gifts<sub>*j*</sub> are easy to give *e<sub>j</sub>* to *e<sub>i</sub>*.  
 b. These gifts<sub>*i*</sub>, these girls<sub>*j*</sub> are easy to give *e<sub>j</sub>* *e<sub>i</sub>*.  
 (Green and Levine 1999: 18)

These cases do not fall under Superiority. They do not involve multiple *wh* questions, and at least some of the nominals are D-linked. Nevertheless, it is tempting to try to assimilate them as another case of hierarchy alignment.

Several things can be seen in these examples. In the first place, unlike the Superiority cases, it does not seem likely that *i*-structure status plays a role here. The sentence-initial nominal can be topical, focal, or interrogative, and either D-linked or not D-linked. Instead, the violin-sonata phenomenon appears to be more strictly syntactic. Second, what is at stake here, as is expressed in the conventional analysis involving nested dependencies, is the order of the gaps. An approach to LDDs that does not recognize the existence of *c*-structure gaps will find it difficult to account for the facts (in particular (39b), where there is no PP). Finally, the examples all involve “Tough Movement” and another LDD.

What distinguishes Tough Movement constructions is that the operator of the LDD construction is the SUBJ of the matrix clause. It is the gap corresponding to this SUBJ that must come first. I propose that this is another manifestation of the alignment of the Relational Hierarchy with linear order, since SUBJ is the most prominent grammatical function on the Relational Hierarchy.

- (42) A gap which corresponds to an element bearing the grammatical function SUBJ must precede another gap in the same clause.

An account such as this is preferable to one which appeals to the nestedness of the dependencies. While it is appealing to attribute the violin-sonata phenomenon to a processing effect which prefers nested dependencies, this cannot be the correct account. As we have seen, there is no general preference for nested dependencies. The manifestation of Superiority in multiple *wh*-fronting languages results in crossing dependencies. And, as shown by Maling and Zaenen (1982), Norwegian and Icelandic allow crossing dependencies:

- (43) Norwegian  
 [Denne gaven her]<sub>i</sub> vil du ikke gjette hvem<sub>j</sub> jeg fikk e<sub>i</sub> fra e<sub>j</sub>.  
 this gift here will you not guess who I got from  
 ‘This gift, you cannot guess who I got from.’
- (44) Icelandic  
 [Þessum krakka hérna]<sub>i</sub>, geturðu aldrei ímyndað þér [hvaða gjöf]<sub>j</sub> ég gaf e<sub>i</sub> e<sub>j</sub>.  
 this boy here can.you never guess there what gift I gave  
 ‘This boy here, you can never guess what gift I gave.’

Like attempts to account for Superiority by appeal to ease of processing, a processing-based analysis of the violin-sonata phenomenon is not justified.

## 4. Why Superiority?

My proposal, then, is that the Superiority Condition is an alignment of linear order with the Relational Hierarchy for non-D-linked *wh* question elements. I will conclude by speculating as to the reason for such a condition. I will first discuss the hierarchy alignment, and then the restriction to non-D-linked elements.

The alignment of distinct hierarchies in syntax has been discussed in the literature primarily from the perspective of Optimality Theory. The application of harmonic alignment is applied to subject choice, split-ergative and differential case marking, and inverse marking by Aissen (1999, 2003), drawing on the original phonology-based proposals by Prince and Smolensky (1993). Even independently of OT, it is fairly clear that language is designed in such a way that prominences on different hierarchies reinforce each other: for example, Agents map to SUBJ, SUBJS are often topical, SUBJ and TOPIC both come early in the sentence. Like other cases of redundancy, such as coarticulation in phonetics, this makes language easier to process. The Superiority Condition is, under this view, not an arbitrary condition, but rather a grammatical condition designed to assist the hearer to decipher the linguistic signal.

The restriction to non-D-linked elements is also not mysterious. D-linking is itself an aid to interpretation; recall that Arnon, Snider, Hofmeister, Jaeger, and Sag (to appear) and Hofmeister, Jaeger, Sag, Arnon, and Snider (2006), from the perspective of processing, refer to D-linked elements as “accessible.” It is the non-D-linked “inaccessible” interrogative elements, which represent questioned new information, which are most in need of reinforcement for processing.

Syntax regulates the way sentences are assembled to express complex concepts. In doing so, it includes conditions that make the concepts easier to understand. For example, it has been argued in Falk (2009) that island constraints in the syntax prevent the creation of structures that are problematic from the perspective of pragmatics and/or processing. Superiority is part of this general function of the syntactic component of the grammar.

## References

- Aissen, Judith L. (1999) “Markedness and Subject Choice in Optimality Theory.” *Natural Language and Linguistic Theory* 17: 673–711.
- Aissen, Judith L. (2003) “Differential Object Marking: Iconicity Vs. Economy.” *Natural Language and Linguistic Theory* 21: 435–83.
- Arnon, Inbal, Neal Snider, Philip Hofmeister, T. Florian Jaeger, and Ivan A. Sag (to appear)

- “Cross-Linguistic Variation in a Processing Account: The Case of Multiple *Wh*-Questions.” *Proceedings of BLS* 32. <http://lingo.stanford.edu/sag/papers/bls-06.pdf>.
- Baker, C. L. (1977) “Comments on the Paper by Culicover and Wexler.” in Peter W. Culicover, Thomas Wasow, and Adrian Akmajian, eds., *Formal Syntax*. 61–70.
- Bolinger, Dwight (1978) “Asking More Than One Thing at a Time.” in Henry Hiz, ed., *Questions*. Dordrecht: D. Reidel. 107–50.
- Bošković, Željko (1997) “On Certain Violations of the Superiority Condition, AgrO, and Economy of Derivation.” *Journal of Linguistics* 33: 227–54.
- Bošković, Željko (2002) “On Multiple *Wh*-Fronting.” *Linguistic Inquiry* 33: 351–83.
- Bresnan, Joan and Jonni M. Kanerva (1989) “Locative Inversion in Chicheŵa: A Case Study of Factorization in Grammar.” *Linguistic Inquiry* 20: 1–50.
- Butt, Miriam and Tracy Holloway King (2000) “Null Elements in Discourse Structure.” in Karumuri Venkata Subbarao, ed., *Papers from the NULLS Seminar*. Delhi: Motilal Banarasidas. [ling.sprachwiss.uni-konstanz.de/pages/home/butt/main/papers/nulls97.ps](http://ling.sprachwiss.uni-konstanz.de/pages/home/butt/main/papers/nulls97.ps).
- Chomsky, Noam (1973) “Conditions on Transformations.” in Stephen R. Anderson and Paul Kiparsky, eds., *A Festschrift for Morris Halle*. New York: Holt Rinehart and Winston. 232–86.
- Chomsky, Noam (1981) *Lectures on Government and Binding*. Dordrecht: Foris.
- Chomsky, Noam (1995) *The Minimalist Program*. Cambridge, Mass.: MIT Press.
- Culicover, Peter W. and Kenneth Wexler (1977) “Some Syntactic Implications of a Theory of Language Learnability.” in Peter W. Culicover, Thomas Wasow, and Adrian Akmajian, eds., *Formal Syntax*. 7–60.
- Falk, Yehuda N. (2007) “Do We Wanna (or Hafta) Have Empty Categories?” in Miriam Butt and Tracy Holloway King, eds., *Proceedings of LFG07*. On-line: CSLI Publication. 184–97. <http://csli-publications.stanford.edu/LFG/12/lfg07.html>.
- Falk, Yehuda N. (2009) “Islands: A Mixed Analysis.” in Miriam Butt and Tracy Holloway King, eds., *Proceedings of LFG09*. On-line: CSLI Publications. 261–81. <http://csli-publications.stanford.edu/LFG/14/lfg09.html>.
- Fodor, Janet Dean (1978) “Parsing Strategies and Constraints on Transformations.” *Linguistic Inquiry* 9: 427–73.
- Ginzburg, Jonathan and Ivan A. Sag (2000) *Interrogative Investigations: The Form, Meaning, and Use of English Interrogatives*. Stanford, Calif.: CSLI Publications.
- Green, Georgia M. and Robert D. Levine (1999) “Introduction.” in Robert D. Levine and Georgia M. Green, eds., *Studies in Contemporary Phrase Structure Grammar*. Cambridge: Cambridge University Press. 1–38.
- Hendrick, Randall and Michael Rochemont (1988) “Complementation, Multiple *wh* and Echo Questions.” in Peter Avery, ed., *Toronto Working Papers in Linguistics, Vol. 9*. 81–97.
- Hofmeister, Philip, T. Florian Jaeger, Ivan A. Sag, Inbal Arnon, and Neal Snider (2006) “Locality and Accessibility in *Wh*-Questions.” *International Conference on Linguistic Evidence, 2006, Feb. 1-4, University of Tübingen, Germany*. <http://lingo.stanford.edu/sag/papers/LE06.pdf>.
- Jaeger, T. Florian and Veronica A. Gerassimova (2002) “Bulgarian Word Order and the Role of the Direct Object Clitic in LFG.” in Miriam Butt and Tracy Holloway King, eds., *Proceedings of the LFG02 Conference, National Technical University of Athens, Athens*. On-line: CSLI Publications. 197–219. <http://csli-publications.stanford.edu/LFG/7/lfg02.html>.
- Kayne, Richard S. (1983) “Connectedness.” *Linguistic Inquiry* 14: 223–49.
- Kuno, Susumu and Jane J. Robinson (1972) “Multiple *Wh* Questions.” *Linguistic Inquiry* 3: 463–87.
- Maling, Joan and Annie Zaenen (1982) “A Phrase Structure Account of Scandinavian Extraction Phenomena.” in Pauline Jacobson and Geoffrey K. Pullum, eds., *The Nature of Syntactic*

- Representation*. 229.
- Mycock, Louise (2006) *A Typology of Constituent Questions: A Lexical-Functional Grammar Analysis of 'Wh'-Questions*. Ph. D. dissertation, University of Manchester.
- Pesetsky, David (1987) "Wh-In-Situ: Movement and Unselective Binding." in Eric J. Reuland and Alice G. B. ter Meulen, eds., *The Representation of (In)definiteness*. Cambridge, Mass.: MIT Press. 98–129.
- Pesetsky, David (2000) *Phrasal Movement and Its Kin*. Cambridge, Mass: MIT Press.
- Prince, Alan and Paul Smolensky (1993) *Optimality Theory: Constraint Interaction in Generative Grammar*. RuCCS Technical Report No. 2, Rutgers University Center for Cognitive Science, Piscataway, NJ.  
<http://roa.rutgers.edu/files/537-0802/537-0802-PRINCE-0-0.PDF>.
- Sag, Ivan A. and Janet Dean Fodor (1994) "Extraction Without Traces." in Raul Aranovich, William Byrne, Susanne Preuss and Martha Senturia, eds., *Proceedings of the Thirteenth West Coast Conference on Formal Linguistics*. Stanford, Calif.: CSLI Publication. 365–84.
- Zubizarreta, Maria Luisa (1998) *Prosody, Focus, and Word Order*. Cambridge, Mass: MIT Press.