The role of linear order in the computation of referential dependencies

Peter W. Culicover

Department of Linguistics, The Ohio State University, 222 Oxley Hall, 1712 Neil Avenue, Columbus, OH 43210-1291, USA

Received 14 June 2011; received in revised form 26 March 2013; accepted 19 July 2013
Available online 27 August 2013

Abstract

I consider several types of referential dependencies: those between bound pronouns and their antecedents, e.g. weak crossover and the classical binding conditions. Constraints on such dependencies have typically been formulated in linguistic theory in terms of conditions on the syntactic structures in which these elements are arrayed. I suggest that a relevant factor in determining the well-formedness of such dependencies is the linear order in which the elements appear. On this view, computation of a referential dependency is sensitive to the extent to which the antecedent is accessible in discourse at the point in the string where the dependent element is processed in the course of interpretation. The evidence suggests that the interaction between linear order and discourse structure provides a more satisfactory account of certain well-known effects than do constraints formulated in terms of syntactic structure. © 2013 Elsevier B.V. All rights reserved.

Keywords: Binding; Anaphora; Weak crossover; Linear order; Discourse representation; Accessibility

1. The linear order hypothesis

A prominent view in classical syntactic theory has been that linear order is an artifact or byproduct of configuration. This view is articulated most clearly by Kayne (1994), who writes: “the human language faculty is in fact rigidly inflexible when it comes to the relation between hierarchical structure and linear order” (xiii) and “phrase structure in fact always completely determines linear order” (3).

I argue here that, to the contrary, linear order plays a central role in determining the acceptability of an array of grammatical phenomena that have previously been accounted for in terms of syntactic configuration, at least in the mainstream literature. I use the term ‘acceptability’ here because the shift to linear order suggests that the correct account does not have to do with grammatical well-formedness per se. Rather, it concerns the complexity of sentence processing and the dynamical construction of discourse representations as this processing proceeds. Specifically, I suggest that as sentence processing proceeds in time, a discourse representation is built whose content and structure are determined by the available syntactic and semantic evidence at each point in the process. The state of the discourse structure at the point at which a pronoun or other dependent expression is encountered plays a role in determining the extent to which the dependent expression can have a particular discourse entity as its antecedent, or how easy it is to identify a particular discourse entity as the antecedent.

The facts to be discussed here suggest that linear order is central to the correct approach for phenomena such as weak crossover, reconstruction and anaphor binding. While the general perspective that I propose is by no means novel...
(see for example Shan and Barker, 2006; Sag et al., 2007), it has not to my knowledge been applied to the analysis of pronominal dependency.\(^1\)

To set the stage, it is useful to first note briefly some key points in the history of the weak crossover phenomenon (WCO), as exemplified in (1a). The asterisk here represents the standard judgment of generative grammarians that (1a) is not possible on the reading where who and his corefer. Although I use this notation to indicate here that a particular coinindexing is unacceptance, I explicitly am not presuming that the unacceptable is a matter of grammar per se. Subscripts are used to mark intended referential dependency, while \(t\) marks the extraction site of the wh-phrase.

(1) a. \(^*\)Who\(_i\) did his\(_i\) mother offend \(t\)?
   b. Who\(_i\) offended his\(_i\) mother?

The term ‘crossover’ was introduced by Postal (1971). ‘Weak’ crossover (1a) was distinguished by Wasow (1972) from ‘strong’ crossover (2a), which is significantly less acceptable on the reading where the wh-phrase and the pronoun are intended to corefer.

(2) a. \(^*\)Who\(_i\) did he\(_i\) say Sandy likes \(t\) best?
   b. Who\(_i\) said he\(_i\) likes Sandy best?

The primary empirical basis for the judgments discussed in this paper is the forty-year literature on crossover phenomena, beginning with Postal (1971) and Wasow (1972). While the precise degree of acceptability of individual examples can and should be systematically investigated, if possible, the broad differences in acceptability between cases of strong crossover, weak crossover and fully acceptable non-crossover appear to be robust and consistent.

On the familiar configurational approach, the unacceptability of (1a) is taken to be a matter of grammaticality. The explanation is typically couched in terms of the syntactic binding relation between the clause-initial wh-phrase, the pronoun, and the gap associated with the canonical position of the wh-phrase, marked here as the trace \(t\). Binding is conventionally determined by the syntactic relation c-command.

(3) a. A constituent c-commands its sister(s) and everything dominated by its sister(s).
   b. A constituent \(\alpha\) binds a constituent \(\beta\) if and only if \(\alpha\) and \(\beta\) are coindexed, and \(\alpha\) c-commands \(\beta\).

In (1), assuming the syntactic structure in (4), who c-commands and is coindexed with its trace \(t\), and the pronoun his\(_i\) in both the WCO example and the well-formed example.

(4) a. 

\[
[s \text{ who}_i \text{ did } [s [np \text{ his}_i \text{ mother}] [vp \text{ offend } t]]]
\]

b. 

\[
[s \text{ who}_i [vp \text{ offended [np his}_i \text{ mother]]}]
\]

Weak crossover has thus been formulated in the mainstream syntax literature as a constraint against the simultaneous (syntactic) binding of a pronoun and a trace, where the trace does not itself c-command the pronoun, as in (5).\(^2\)

(5) Who\(_i\) (did you say) \(t\); offended his\(_i\) mother.

Such a formulation is the Bijection Principle of Koopman and Sportiche (1982): “An operator can bind at most one variable.”

From this formulation it is a simple step to conclude that in all examples that show a similar binding violation, there is a similar configuration in which the antecedent c-commands the pronoun and a trace.\(^3\) This is the mainstream explanation for sentences such as (6), where the quantifier phrase follows the pronoun in the observed constituent order.

(6) \(^*\)His\(_i\) mother offended no boy,

---

1. Kempson and Meyer-Viol (2002) propose, as best as I can tell, an approach to grammar in which the grammar itself is incremental. My focus here is on the dynamics of the interpretative mechanisms, and I make no claims about the architecture of the grammar itself.

2. If we assume that the structure of (1b), is

\[(i) [s \text{ who}_i [t [vp \text{ offended [np his}_i \text{ mother]]}]]\]

as many generative grammarians do, the formulation of the constraint must be adjusted so that the wh-phrase can c-command a pronoun and a trace only if the trace c-commands the pronoun.

3. Such a generalization step is an instance of a methodology that is referred to as Uniformity in Culicover and Jackendoff (2005). As pointed out there, it is a methodological approach that is justified only as long as the consequences are valid. Culicover and Jackendoff argue that in many, if not most cases, it is not.
On classical analyses of this type of violation there is invisible movement of the quantifier phrase so that it appears in the same configuration as the moved wh-phrase at LF.

(7) [no boy], [his mother offended t1]

As a consequence of this movement, the configuration in (7) is the essentially the same as that of (1a).

At the same time, it is clear that in the cases where binding of the pronoun by the quantifier phrase (including the wh-phrase) is problematic, the pronoun precedes the position in which the quantifier phrase acquires its grammatical function and interpretation with respect to the clause that contains it. In fact, Chomsky (1976) proposed the following condition: “A variable cannot be the antecedent of a pronoun to its left.” This is the Leftness Condition.

It is conceivable that the Leftness Condition is simply a superficial reflection of a deeper configurational property of English. For example, it has been proposed that a pronoun cannot be interpreted as a variable unless it is c-commanded and coindexed with an argument – see for example Reinhart (1976, 1983). Since English is for the most part right branching, a pronoun that precedes its antecedent will not be c-commanded by the antecedent, and on this definition, it is not interpretable as a variable. Such an account of course leaves open the question of why a pronoun must be c-commanded by an argument in order to be bound by that argument.

Furthermore, Wasow (1972:163) shows that many examples whose syntactic configuration classifies them as ‘weak crossover violations’ are in fact quite acceptable. And Wasow (2002:158) highlights the fact that WCO violations vary significantly in acceptability, a phenomenon that should be accounted for but that has been largely ignored in the mainstream literature.

The fact that one can construct a strictly syntactic account of the core cases of WCO does not mean that such an account is correct. In the present case we face some challenging questions regarding such an account. Most fundamentally, why should something like the Bijection Principle hold at all, if it does? And if the Bijection Principle is not sustainable in the end, because it does not cover all of the relevant cases, why should it nevertheless appear that something like the Leftness Condition holds? And if Leftness turns out to be a superior characterization of the phenomena, why is linear order relevant to whether a particular binding relation may hold? And why do WCO violations vary in acceptability?

The answers to these questions are related, and comprise the Linear Order Hypothesis. Crucially, I assume that the computational requirements of constructing a relation between discourse entities correlates with the perceived acceptability of the sentence under the intended interpretation. If there is difficulty in identifying the antecedent at a particular point, but not total failure, the judgment is a graded, not a categorical one. I suggest that these graded acceptability judgments may correspond to statements in the linguistics literature to the effect that one or another configuration is ‘ungrammatical’ or ‘somewhat ungrammatical’ or ‘less than fully grammatical’.

I therefore propose to take WCO and similar phenomena out of the domain of grammar per se, and put it into the domain of processing. This idea, while not the standard one in linguistic theory, is not without precedent. It is found in the seminal work of Kluender (1992, 1999, 2004), Kluender and Kutas (1993) and, in a somewhat different form, in that of Hawkins (1994, 1999, 2004). The idea that the processing of discourse is central to understanding how such phenomena work is due primarily to Erteschik (1973) and Erteschik-Shir (1997, 2006, 2007). Recent work by Philip Hofmeister, Ivan Sag and colleagues at Stanford on Superiority and other putative grammatical constraints supports and extends this perspective (Am non et al., 2006; Hofmeister et al., 2007; Sag et al., 2007).

The organization of the remainder of this article is as follows. In section 2 I summarize the main non-configurational factors that affect the acceptability of WCO. In section 3 I extend the account to interactions between binding and chains, that is, reconstruction and connectivity, and consider a number of other cases involving binding relations that are sensitive to the linear order of the antecedent and the dependent element. Section 4 is concerned with the plausibility of the proposal that what is at work in determining WCO judgments is the processing of discourse in time. Section 5 summarizes and points toward further research with this general perspective.

2. Weak crossover

In this section I summarize the main non-configurational factors that contribute to variable WCO effects. Briefly, it appears that WCO occurs for two reasons, both of which have to do with the interpretation of a pronoun as dependent on an quantified NP antecedent. (i) the antecedent is not thematically interpreted at the point at which the pronoun is
interpreted, and (ii) there is insufficient information about the antecedent to allow the pronoun to be linked to the discourse representation.

I suggest both of these factors contribute to the ‘accessibility’ of the antecedent in the discourse. In cases of less accessibility, there is greater processing effort in linking the pronoun to the antecedent. This processing effort is experienced subjectively as ill-formedness. Such judgments form the basis for treating these cases as ‘ungrammatical’ and for accounting for them in terms of grammatical principles.

The discussion proceeds as follows. Section 2.1 demonstrates that given the syntactic configuration associated with WCO, the crucial factor that determines whether or not there is a WCO effect is the relative order of the gap corresponding to the wh-phrase and the pronoun bound by the wh-phrase. Section 2.2 shows that WCO effects in reversible pseudo-clefts with the WCO configuration are sensitive to the linear order of the focus and the embedded question, suggesting again that what is critical is not the structure but the order. Section 2.3 shows that given the WCO configuration, the WCO effect is ameliorated when the wh-phrase is made more specific, and thereby more discourse accessible. This constitutes further evidence that the source of WCO effects is not a configurational one, but has to do with interpretation of the pronoun at the point in the sentence where it is encountered in the course of processing.

2.1. If the gap corresponding to the wh-phrase precedes the pronoun, there is no WCO violation

The first group of cases consists of linear order asymmetries, where only the relative order of the gap corresponding to the extracted wh-phrase with respect to the pronoun varies and there is a corresponding difference in acceptability. The binding relations, determined by the syntactic structure, appear to be identical, contrary to what we would expect given a configurational WCO constraint.

Wasow (1979:157) cites the examples in (8).

(8)  
(a) Who, did Mary talk to $t_1$ about his, sister?  
(b) *Who, did Mary talk about his, sister to $t_1$?

Bresnan (1995:255), citing similar observations of Jackendoff (1990a), gives the examples in (9)–(10).

(9)  
(a) *Who, did you talk about her, parents with $t_1$?  
(b) Who, did you talk with $t_1$ about her, parents?

(10)  
(a) *Who, did you talk with her, parents about $t_1$?  
(b) Who, did you talk about $t_1$ with her, parents?

Because of the branching of the PPs in these examples, neither the trace nor the pronoun c-commands the other. Hence in every case there is a wh-phrase that binds a pronoun and a trace, where neither binds the other, a configuration that violates the Bijection Principle.

The key here is that in the acceptable b-examples, the gap associated with who is ordered before her. Assuming that the interpretation is constructed dynamically in the course of processing the string, the thematic role of the wh-phrase is established before the pronoun is encountered. The assignment of a thematic role appears to be a factor that makes a discourse entity more accessible to subsequent linking to a pronoun.

Note the following similar examples with full NPs in the position of the extracted wh-phrase show the same difference in acceptability.

---

7 It is possible, of course, to stipulate by brute force that the complement of a preposition c-commands to its right, but such a move falls well short of an explanation of what is going on in such cases. Moreover, it is possible to construct cases where the trace is further embedded in the PP. In such cases, the linear order in which the trace follows the pronoun is less acceptable (a WCO effect) than when the trace precedes the pronoun. For example, consider (i.b).

(i)  
(a) *Who, did you work on her, project with some friends of $t_1$?  
(b) Who, did you work with some friends of $t_1$ on her, project?

It cannot be plausibly argued that the trace in (i.b) c-commands the pronoun.

8 While examples of the form (9b) are not common, they do occur. A search of Google found 5 examples of $N_1$ who she talked with about her $N_2$, 6 of $N_1$ who he talked with about his $N_2$. On the other hand, the search found no examples for the form $N_1$ who she talked about her $N_2$ with $N_1$ who he talked about his $N_2$ with. No examples were found of the form (10b). This latter case appears to be one of those where additional forms of evidence need to be brought to bear to more firmly establish relative acceptability.

9 Thanks to a reviewer for raising this point.
(11) a. I talked about her father to Cordelia.
b. I talked to Cordelia about her father.
c. I talked to her father about Cordelia.
d. I talked about Cordelia to her father.

These examples suggest that computing the reference of the pronoun is facilitated when the antecedent precedes it. The difference in the putative WCO examples in (9)–(10) is that the antecedent precedes the dependent in all of them. The main point is that those examples where the antecedent has a thematic role at the point at which the pronoun is encountered in the string are more acceptable.

A similar but different type of case involves across-the-board extraction in conjointed structures. As noted by Munn (2001:374), when the WCO configuration occurs in the left conjunct, the consequence is unacceptability, but not when the WCO configuration occurs in the right conjunct. ATB extraction produces two chains that share the same head but have different tails. In both (12a) and (12b) there is one chain that crosses over a pronoun. The pronoun and the trace do not c-command one another, so there should be a violation of the Bijection Principle and a WCO effect. (As in other examples, my use of the conventional ‘?’ and ‘??’ here and throughout is intended to indicate degrees of unacceptability, not strict ungrammaticality in the narrow sense.)

(12) a. ?Who does his mother love t and Mary dislike t?
b. Who i does Mary dislike t and his mother love t?
(13) a. ?I found out who, his mother loves t and Mary dislikes t.
b. I found out who, Mary dislikes t and his mother loves t.

A similar pattern is found in parasitic gap constructions.

(14) a. ?Who, did his mother pass around a picture of t, without Mary actually ever meeting pg?
b. Who i did Mary pass around a picture of t, without his mother actually ever meeting pg?

(15) a. ?I found out who, his mother passed around a picture of t, without Mary ever meeting pg?
b. I found out who, Mary passed around a picture of t, without his mother ever meeting pg?

In every case, the wh-phrase precedes the pronoun. Again, the generalization here appears to be that if the wh-phrase gets a thematic interpretation from the trace before the pronoun is encountered, it is easier to interpret the pronoun as dependent on it. Most strikingly, the corresponding examples with quantifiers in situ do not produce equivalently acceptable interpretations.

(16) a. ??His mother loves every boy and Mary dislikes him.
b. ??Mary dislikes him and his mother dislikes every boy.

These examples suggest, again, that a determining factor in whether there is a WCO violation is the linear order in which the operator appears with respect to the pronoun.

2.2. Reversing a pseudo-cleft determines whether there is a WCO violation or not

Another group of cases where linear order appears to play a role consists of reversed pseudo-clefts, where the argument structure and the logical form are identical, but the order in which the arguments appear varies. A typical pair is the following.

(17) a. ??The person who, his mother loves t, most is no one.
b. No one is the person who, his mother loves t, most.

If the relative pronoun is not an operator, there should be no WCO configuration in either example, since there is no operator that binds both the pronoun and the trace. Hence both examples should be acceptable. On the other hand, if we assume that the relative pronoun is an operator, both examples should be WCO configurations and be unacceptable. Since the only difference between these examples is the order in which the pronoun his and the quantifier no one appear in the string, it again appears that linear order alone is able to ameliorate WCO.

Notice that in the examples in (17) the antecedent is quantificational. It appears that the referentiality of the NP affects acceptability here as it does in other WCO cases. If we replace no one with John, the difference in acceptability is substantially diminished as seen in (18).
(18) a. The person who, his, mother loves $t_i$ most is John$_i$.
b. John$_i$ is the person who his, mother loves $t_i$ most.

2.3. More specificity of wh-phrase improves WCO

The major paradox of WCO configurations is that they do not always produce ill-formedness judgments. The extent of the paradox from a theory-internal perspective depends on the exact formulation of the condition. In the original formulations, WCO violations were ruled out in terms of the syntactic configuration — the pronoun is assumed to be a (syntactic) variable (syntactically) bound by the wh-phrase or quantifier phrase in an A’ position, possibly in an invisible syntactic representation, such as LF. However, certain later formulations of the condition, e.g. Lasnik and Stowell (1991), formulated the condition in terms of the binding of the pronoun by any A’ binder, whether or not it is quantificational. This more general formulation produces a class of apparent counterexamples to WCO, called ‘weakest crossover’, e.g.,

(19) a. John$_i$, his, mother loves $t_i$.
b. The man, who his, mother loves $t_i$ the most is John.

Lasnik and Stowell argued that such cases fail to violate the WCO condition because they do not in fact contain a trace, but an empty epithet, in the position marked by $t_i$ in (19).  

However, Wasow (1979:163) observes that WCO violations are more natural when the wh-phrase contains which or how many, which make them more ‘determinative’ in his terms. E.g.,

(20) a. Which picture, did the man who painted it, refuse to sell $t_i$?
b. How many dachshund$_i$, does your friend who breeds them own $t_i$?

(21) a. Which of John’s teachers, do the people who know them, all respect most $t_i$?
b. How many of the demonstrators, did the police who arrested them, beat $t_i$ up?

A reviewer points out that in Wasow’s examples the pronoun is in a relative clause, and suggests that this might be a consequence of late insertion of the relative clause, following a proposal of Lebeaux (1988). While such an account cannot be ruled out absolutely, it does present some challenges. On the classical account, the Bijection principle applies at LF, after the movement of operators to A’ position. In order to prevent the principle from applying to pronouns in relative clauses, late insertion of the relative clause would have to follow the mapping from syntactic structure to LF and the movement of operators. But then it is not clear how to get the interpretation of so-called ‘donkey’ sentences.

(22) Every farmer who owns a donkey beats it.

---

10 It is in fact possible to find naturally occurring acceptable WCO violations. There are 120 examples in the Corpus of American Contemporary English (COCA) that satisfy the search criterion n-pronoun+wh-pronoun$_\text{poss}$, exemplified by (19b). The examples that appear to be WCO violations are the following.

(i) a. . . will play on HBO later this year) presents a complex portrait of a man who his biographer, the Pulitzer Prize-winning political historian
Samantha Power, describes in . . .
b. Have you ever thought as to why people are so vituperative about this president who his critics say doesn’t have a philosophy – you can’t call him liberal . . .
c. Wojciechowski has become a symbol for Duke’s revival, the gritty little guy whom his fellow students bark at in appreciation for his
dogged defensive style.
d. some people are very calm about it, and then I’ve had gate agents who their faces get very red, and you can tell that they’re very comfortable . . .

Only the first three are weak crossover violations; (i,d) is a resumptive pronoun. A search for N,-pronoun$_\text{wh}$,-pronoun$_\text{poss}$ was satisfied by examples in (ii).

(ii) a. John, Jr., whom his family called John John, seldom talked with reporters about himself or his memories . . .
b. The farmer, whom his parents had located to hide them, was a Christian.
c. Gary has remarried, and Greg and his wife have had two children, whom their grandfather barely knows.
d. It might be that a sluggish bond-servant, or an undutiful child, whom his parents had given over to the civil authority, was to be corrected at . . .

By way of comparison, there are no WCO violations in the corpus consisting of questions of the form pronoun$_\text{wh}$,-pronoun$_\text{poss}$ . . . and there are no questions of the form which-N,-n$_\text{aux}$,-pronoun$_\text{poss}$ . . . WCO violations or otherwise. Nor are there any WCO in embedded questions of the form which-n-pronoun$_\text{poss}$ . . .

11 Wasow marks the examples in (20) with ‘?’ and those in (21) as fully acceptable.
In such a sentence, *it* is interpreted as a variable only in virtue of the binding of a *donkey* by *every*. But if the relative clause is not inserted until LF and after binding, it would be necessary to carry out a re-binding computation.

At the same time, there are plenty of acceptable sentences in which the pronoun appears in a complement to the noun, suggesting that the relative clause is not the source of the acceptability. I find no discernible difference between the examples in (23)–(24), which lack relative clauses, and those in (20)–(21).

(23) a. Which picture, did the man with a photograph of it, refuse to sell *it*?
    b. How many dachshunds, does your friend from their, original breeding farm own *it*?
(24) a. Which of John’s teachers, do the people from their, alma mater all respect most *it*?
    b. How many of the demonstrators, did the police with high resolution pictures of them, beat *it* up?

Along related lines, Levine and Hukari (2008) point out that the failure of such examples to violate the conditions that define WCO is due not to the formal properties of the gap, but rather to the fact that Lasnik and Stowell’s formulation of the conditions on WCO is too broad – WCO should be restricted to quantificational antecedents, and not generalized to all *A* binders.

Levine and Hukari observe, furthermore, that not all WCO configurations are equally unacceptable, even when the antecedent is a interrogative wh-phrase. They suggest that the unacceptability correlates with the degree of ‘referential support’ that the antecedent provides the pronoun. The parallels with Wasow’s examples are striking. E.g.,

(25) a. *?Who, did his, dean publicly denounce *it*?
    b. ??Which professor, did his, dean publicly denounce *it*?
    c. {?[Which distinguished molecular biologist that I used to work with]i, did his, dean publicly denounce *it*?}

From their discussion, it appears that referential support correlates to some extent with a definiteness presupposition (hence the improvement with *which*) and to some extent with descriptive content (hence the improvement when a relative clause is added to the description). While the difference between (25a,b) and between (25b,c) is difficult to pinpoint, and none of the examples are particularly natural, the difference between (25a,c) appears to be quite sharp.

One way to construe these observations is that as an antecedent becomes more ‘referential’, it is no longer interpreted as an operator, and therefore does not fall under whatever is responsible for WCO effects. The plausibility of this interpretation of the facts is supported by the observation that a wh-phrase can be the discourse antecedent of a pronoun if it is more specific, as shown by the following examples.

(26) a. Ok, whoi, wants to go to the zoo? *Has he, been behaving himself?*
    b. Which kidi, wants to go to the zoo? *?Has he, been behaving himself?*
    c. Which very special and intelligent studenti, wants to go to the zoo? *Has he, been behaving himself?*

Example (26a) is intended to be interpreted with *who* as quantificational, ranging over an otherwise unspecified but contextually defined set of individuals. In (26b) there is a set of kids that *which kid* is drawn from, while in (26c) the person asking the question is plausibly assumed to have a specific student in mind. Again, the difference between (26a) and (26c) seems to me to be a sharp one.

Similarly Falco (2007) offers an account of WCO in terms of the ‘specificity’ of the *A* binder. Falco suggests that the acceptability of non-quantificational WCO (weakest crossover in Lasnik and Stowell’s terms) is due to the fact that the non-quantificational antecedent is sufficiently specific. On this account, the acceptability of weakest crossover and of examples such as (25b,c) fall under the same formulation – the *A* binders are non-quantificational and thus can serve as antecedents of the following pronouns.

---

12 The judgments are theirs, I have changed the pronoun in the original examples to *his* in order to remove the potential effect of the non-stereotypical *her*.

13 In an unpublished study, Wasow and Clausen (2011) found experimental evidence to support the claim that WCO is less acceptable than the corresponding non-crossover construction, and that more informative wh-antecedents produce greater acceptability than do simple wh-antecedents.

14 This is not to suggest that referentiality is the only factor that bears on the acceptability of WCO. A reviewer cites the following examples.

(i) a. *whoid she, boss* (but only her, boss) regard as a priceless asset to the company?
    b. who did *her* boss regard as a priceless asset to the company but everyone else detest?
    I agree that (i,a) is relatively good, but (i,b) strikes me as marginal.
It should be noted that the role of specificity in the acceptability of WCO bears a striking resemblance to what has come to be called ‘D-linking’ in the domain of superiority effects (Pesetsky, 1987). As the examples in (27) show, wh-phrases of the form which $N$ in both $A'$ position and in situ ameliorate the superiority violation. We see the most amelioration when which $N$ is in both positions, and modest amelioration when there is only one which $N$.

(27) a. *What, did who write $t$?
   b. ??Which paper, did who write $t$?
   c. ?What, did which student write $t$?
   d. Which paper, did which student write $t$?

Related to this phenomenon is work by Hofmeister (2007). In a study of this and other cases, Hofmeister finds evidence that ‘informativity’ of a filler correlates with acceptability judgments of filler-gap dependencies. His hypothesis is that the more informative a filler is, the easier it is to retrieve it from memory in order to form the chain. And Sag et al. (2008) propose a processing account of superiority effects along these lines.

3. Order and binding

In this section I consider several additional phenomena that suggest that linear order is relevant to the interpretation of binding relations. I do not claim, however, that binding can be reduced entirely to linear order – the relevance of structure is clear. But whether the structure that is relevant is syntactic structure per se is less clear. I present evidence that shows that when syntactic configuration is varied but conceptual structure is held constant, it is the latter that accounts for the possible binding relations, subject to the constraints imposed by linear order.

3.1. The binding conditions

In linguistic theory, the relationship between a pronoun and its antecedent are typically accounted for in terms of some variant of the binding conditions summarized in (28), where ‘bound’ is understood as in (29).\(^\text{15}\)

(28) Binding Conditions.
   A. A reflexive must be locally bound.
   B. A (non-reflexive) pronoun must be locally free (not bound).
   C. A pronoun cannot bind its antecedent.

(29) $\alpha$ binds $\beta$ if $\alpha$ and $\beta$ are coindexed and $\alpha$ c-commands $\beta$.

Understood syntactically, condition A says in effect that the antecedent of a reflexive must be in the same clause as the reflexive and c-command it, with some well-known complications concerning ‘raising’ constructions that we can set aside here (see Chomsky, 1973).

But (30b) shows that when an antecedent is in $A'$ position, it still acts as a local antecedent of a reflexive, and (31b) shows that if a phrase in $A'$ position is not a local antecedent in its canonical position, it is not a local antecedent in the $A'$ position.

(30) a. I don’t think [Mary, is proud of herself].
   b. Mary, I don’t think [it is proud of herself].
(31) a. *I told herself, [Mary, would win].
   b. *Mary, I told herself, [it, would win].

Similarly, condition B says that the antecedent of a pronoun cannot c-command it and be in the same clause. The examples in (32) show that locating the antecedent in a higher $A'$ position does not change the requirement that a pronoun cannot have a local antecedent –

\(^{15}\) The definitions of ‘locally’ and ‘bound/free’ are in fact at issue, depending on one’s assumptions about syntactic structure, conceptual structure, the relationship between them, and at what level these conditions hold. See Reinhart and Reuland (1993). I return to this matter in section 3.3.
(32) a. *I think [Mary$_i$ is proud of her$_j$].
   b. *Mary$_i$, I think [ti is proud of her$_j$].

The antecedents behave as though they are in the position of the trace as far as these binding conditions are concerned.

On the other hand, movement of a reflexive to an $A'$ position appears to preserve its relationship with its antecedent.

(33) Himself$_i$, Otto$_i$ was most proud of $t_i$.

Similarly, when we move a pronoun, its relationships to potential antecedents are not changed.

(34) a. *Her$_i$, I think [Mary$_i$ is proud of ti]. (cf. (32a))
   b. *Her$_i$, Mary$_i$ loves $t_i$. (cf. *Mary$_i$, loves her$_i$)
(35) a. John$_i$, thinks that Mary$_i$ is proud of him$_i$.
   b. John$_i$, thinks that him$_i$, Mary$_i$ is proud of $t_i$.
   c. Him$_i$, John$_i$ thinks that Mary$_i$ is proud of $t_i$.

It is possible to ensure that the binding conditions apply as we see here in a theory that assumes movement (in this case of him/himself), by assuming either that movement leaves behind an invisible copy of what is moved, that the binding conditions apply after 'reconstruction' of the moved constituent back into the position of its trace, or that they apply prior to the movement. In any case, it appears at first glance that a strictly configurational account of the binding conditions is possible.

However, there are many cases where the binding conditions apply but movement is not a viable analysis. In these cases, it appears that there are in fact conditions on binding, but they are conditions on conceptual structure and not syntactic structure.\footnote{For a proposal that binding is computed on the basis of LF representations, but is sensitive to linear order, see G"{o}bbel (this volume).} And, as seen in the next section, in these cases, linear order also appears to play a role.

3.2. Non-configurational conditions on anaphora

The first piece of evidence that anaphora is governed by non-configurational conditions is the well-known demonstration by Barss and Lasnik (1986) that in the English double object construction, the antecedent must precede a reflexive or reciprocal.

(36) a. Mary showed Bill$_i$, himself$_i$.
   b. *Mary showed himself$_i$, Bill$_i$.
(37) a. Mary showed Bill$_i$, and Susan$_i$ each other$_{i+j}$.
   b. *Mary showed each other$_{i+j}$ Bill$_i$, and Susan$_i$.

Larson (1988) showed how to preserve a configurational account of anaphor binding in the VP, but only by assuming an abstract structure in which the antecedent is higher than the anaphor at the point at which binding condition $A'$ applies. On this analysis, the verb must move from a deeply embedded position in the structure to the observed linear position immediately after the subject. But this analysis is complicated somewhat by the fact that the antecedent may sometimes be the complement of a preposition. Because of the branching structure of PP, the complement strictly speaking does not c-command constituents to the right of it. (38a) shows that the complement of to can bind anaphors that follow it. (39b) shows that the complement of about cannot, suggesting that it is not syntax but conceptual structure and a thematic hierarchy that is relevant (Jackendoff, 1972). (38b) and (39a) show that the antecedent must precede the anaphor.

(38) a. I talked to the students$_i$, about each other$_i$.
   b. *I talked about each other$_i$, to the students$_i$.
(39) a. *I talked to each other$_i$, about the students$_i$.
   b. *I talked about the students$_i$, to each other$_i$.
Hence it is not clear that a c-command condition can be maintained for all cases of binding of anaphora.\footnote{17} Moreover, when there is variable constituent order in VP, it is the first constituent that is the antecedent. In the following, Heavy NP Shift of a direct object makes it ineligible as an antecedent.

(40) a. I described [the students] that I met to each other.
   b. *I described to each other, [the students; that I met].

But Heavy NP Shift of an anaphor allows it to be bound by an antecedent in the PP.

(41) a. *I mentioned [each other;’s projects] to the students.
   b. I described to the students, [each other;’s projects].

Again, in these cases it appears that the configuration is overridden by the linear order.

The simplest generalization that covers all of these cases is that the antecedent must be a co-argument of the anaphor, and precede it.\footnote{18} The co-argument condition applies to anaphors that are arguments of verbs, and excludes those that are embedded in NPs, the so-called ‘exempt anaphors’ (Pollard and Sag, 1992).

To consider the first case, an anaphor in an argument position must be bound by a local antecedent that it agrees with.

(42) \[\{\text{The men} \\
   \quad \text{Mary}\}\] showed himself . . .

At the point at which himself is encountered there is a local antecedent, but it does not match the reflexive in gender. In the absence of any alternatives, condition A would rule out this sequence as ungrammatical. Because of the precedence condition, it doesn’t matter that there is a following potential antecedent that agrees with the reflexive and is a co-argument of it, as seen in (43).

\footnote{17} Pesetsky (1995) shows that it is possible to avoid this consequence by positing a second, parallel syntactic structure for every sentence such that in this second structure, the c-command relation is not affected by the preposition. For example, in (38a), the students c-commands and thus binds each other, and similarly for the other examples. Given that such a structure already exists in the form of an independently motivated conceptual structure, it would appear that Pesetsky’s solution is a syntactization of the semantics, a step often seen in mainstream syntactic theory. Culicover and Jackendoff (2005) argue that to the extent that semantic structure is independently required, Occam’s Razor renders such purely syntactic accounts superfluous.

On the other hand, a reviewer observes that in German, which does not have P-stranding, sentences parallel to the grammatical (38a) are not possible.

(i) a. *Ich sprach mit Hans über sich.
   I spoke with Hans about self
   ‘I spoke with Hans about himself.’
   b. *Ich warnte vor den Kollegen einander.
   I warned from the colleagues each other
   ‘I warned each other about the colleagues.’

This difference between English and German is an interesting one, and warrants a fuller consideration than I can devote to it here. In any case, I don’t think that it can be attributed to a restructuring of P with V. P-stranding is fully general in English, to the point that it occurs in NPs, APs, and in VPs with arguments and adjuncts separating the V and the P. So, for example, we get binding in cases such as (ii).

(ii) I gave a little present at the party to every child, just as she, was getting ready to leave.

In order to get every child to c-command the pronoun, we would have to adjoin the preposition to gave a little present at the party, a clearly ad hoc solution that lacks any independent syntactic motivation.

Moreover, it is important to observe that there are cases that permit P-stranding but not binding. P-stranding is possible with about (Who did you talk about?) but (39b), where the antecedent is the complement of about and precedes the anaphor is not a well-formed case of binding.

In the cases considered here, ‘co-argument’ is equivalent to ‘argument of the same conceptual structure relation’. However, there are more complex cases where this definition does not work, e.g. where the direct object of a verb does not bear a thematic role with respect to that verb, as in (i).

(i) Mary, expects herself, to be the nominee.

The notion ‘co-argument’ must in fact cover cases where the anaphor and the antecedent bear grammatical functions within the same syntactic domain, but are not semantically related; see Pollard and Sag (1994) and Culicover and Jackendoff (2005) for discussion. In many respects this appears to be a version of Langacker’s original account of pronominal coreference (Langacker, 1969), which was formulated in terms of precede and command, the latter meaning essentially ‘same S’.
(43) \( \{ \text{The men}, \text{Mary} \} \) showed himself \ldots to John.

On the other hand, suppose that the reflexive is an exempt anaphor, e.g. a reflexive in a phrase such as stories about herself. Unlike anaphors that are arguments of verbs, exempt anaphors may have long distance antecedents. So, in (44c, d), a reflexive in the subject of the lower clause can have an antecedent in the higher clause.

(44) a. Mary\(_i\) told John\(_j\) stories about herself\(_i\).
   b. Mary\(_i\) told John\(_j\) stories about himself\(_i\).
   c. Mary\(_i\) told John\(_j\) [that [stories about herself\(_i\)] are flattering].
   d. Mary\(_i\) told John\(_j\) [that [stories about himself\(_i\)] are flattering].

And, an exempt anaphor, like a pronoun, may actually precede its antecedent, as in (45).

(45) Stories about himself, make John\(_i\) nervous.

The examples in (46) show that when a phrase containing an exempt anaphor is in an A’ position, the antecedent of the exempt anaphor may be any potential antecedent that follows it in the same clause and precedes the trace. So (46a) with herself is unacceptable because construction of the chain makes John the first and only potential antecedent, while (46b) is acceptable with either herself or himself because the trace follows both potential antecedents.

(46) a. [Which pictures of ?herself\(_i\)/himself\(_i\)] did John\(_j\) sell \( t_i \) to Mary\(_k\)?
   b. [Which pictures of herself\(_i\)/himself\(_i\)] did John\(_j\) convince Mary\(_k\) to sell \( t_i \)?

The upshot is that there is a structural condition on the binding of anaphors, but not one expressed in terms of syntactic structure, defined strictly in terms of c-command. The condition holds at conceptual structure. Relations between elements in CS are determined by properties of a sentence that correlate with configuration, that is, grammatical function and thematic role (Culicover and Jackendoff, 2005), but are not necessarily isomorphic to configuration. For instance, the complement of a PP may be a CS argument, if the lexical entry of the main verb specifies that it corresponds to a thematic role. Hence binding in examples like I talked to the students about each other is possible, even though there is no syntactic binding relation in terms of c-command. The sentence satisfies the two conditions on binding: (i) the students precedes each other, and (ii) the students binds each other in CS.\(^{19}\)

Recognizing the effects of linear order on binding allows us to account for the Kearney puzzle involving exempt anaphors, exemplified in (47) (Kearney, 1983). (The absence of marking on (47a) and the “*” on (47b) are intended simply to highlight the difference in acceptability between them.)

(47) a. [Which books about himself\(_i\)] did John\(_j\) file \( t_i \) without Mary\(_k\) first looking at pg\(_j\)?
   b. *[Which books about himself\(_i\)] did John\(_j\) file \( t_i \) without Mary\(_k\) first looking at pg\(_j\)?

On the view that which books heads multiple chains, one containing t and one containing pg, both sentences should be ungrammatical.\(^{20}\) This is because in both sentences, the reflexive will be bound by one antecedent with which it agrees (e.g. John in the case of (47a)), and one antecedent with which it does not agree (e.g. Mary in the case of (47b)).

Levine and Hukari (2006) point out that the Kearney paradigm occurs even when it is difficult if not impossible to justify an analysis in which the phrase containing the reflexive, in this case which books about himself/herself, heads an A’ chain containing the parasitic gap. Adapting their examples, consider the following.

(48) a. [Pictures of himself\(_i\)] would be tough for John\(_j\) to file \( t_i \) without Mary\(_k\) first looking at pg\(_j\).
   b. *[Pictures of himself\(_i\)] would be tough for John\(_j\) to file \( t_i \) without Mary\(_k\) first looking at pg\(_j\).

(49) a. [Pictures of himself\(_i\)] would be tough for John\(_j\) to file \( t_i \) without Mary\(_k\) first looking at pg\(_j\).
   b. *[Pictures of himself\(_i\)] would be tough for John\(_j\) to file \( t_i \) without Mary\(_k\) first looking at pg\(_j\).

---

\(^{19}\) I do not have space here to review the nature and function of conceptual structure. For extensive discussion, see Culicover and Jackendoff (2005) and Jackendoff (1976, 1983, 1990b, 1997, 2002).

\(^{20}\) While approaches to parasitic gaps differ on the details, most agree that the pg, like a trace, forms a chain with its antecedent. See, for example, Chomsky (1986), Frampton (1990), Kayne (1984), Munn (2001), Sag (1983) and Williams (1990).
There is no movement of pictures of himself/herself to the subject position of tough (things). What is plausibly going on here is that the reference of the exempt anaphor is left unresolved in the course of processing until a suitable antecedent is encountered. In (48) and (49) this antecedent is John. Since John can be the antecedent of himself but not herself, the b-examples fail at the point at which the gap is encountered.

In the examples in (50), both the true gap and the parasitic gap precede the exempt anaphor. In this case, either John or Mary is available as a candidate for the reflexive, and therefore they are equally acceptable.

(50) What, would be tough(est) for John, to file t, without Mary, first looking at pg, are (these) pictures of \{ himself, herself, \}.

Strikingly, when we reverse the phrases around are, the exempt anaphor again precedes both potential antecedents, and only himself is acceptable, as shown in (51).

(51) These pictures of \{ himself, herself, \} are what, would be tough(est) for John, to file t, without Mary, first looking at pg,.

Again, we have evidence that the computation of a grammatical dependency is sensitive to linear order.

3.3. Connectivity and ordering effects

In cases of 'connectivity', the linear order and the syntactic structure interact to produce mismatches with conceptual structure. The term 'connectivity' covers a range of cases in which a constituent in one position behaves semantically as though it is in another position.

Some of the most striking cases of connectivity involve binding. One type of connectivity (not always recognized as such) is topicalization, where a constituent in an A’ position is understood as satisfying the grammatical functions associated with a gap (see Culicover and Jackendoff, 2005). This relationship extends to the binding of anaphors, as we have seen. In (52), for instance, there is a mismatch between the properties of the reflexive himself and those of Mary, and unacceptability results. This is the case even though the NP John appears in the sentence following the gap.

(52) *Himself, Mary introduced t to John.

So himself behaves as though it is in the position of the gap.\(^{21}\)

Topicalization displays the same effects for condition B-type and C-type violations.\(^{22}\)

(53) a. *Mary, is proud of her.
   b. *Her, Mary, is proud of t.

(54) a. *She, thinks we are going to hire Mary.
   b. *Mary, she, thinks that we are going to hire t.

\(^{21}\) There are well-known cases where the topicalized or clefted reflexive takes an antecedent that is too far from the gap (Pinkham and Hankamer, 1975). E.g.,

(i) a. Herself, Mary, thinks that John, insulted t, just to be mean.
   b. It is herself, that Mary, thinks that John, insulted t, just to be mean.

but

(ii) *Mary, thinks that John, talked about herself, just to be irritating.

The reflexive may also be linked to the gap, in which case a lower antecedent is selected.

(iii) a. Herself, John, thinks that Mary, talked about t, just to be irritating.
   b. It is herself, that John, thinks that Mary, talked about t, just to be irritating.

One way to account for the topicalization/cleft cases is to resolve the antecedent of the reflexive before the gap is encountered. Space limitations preclude working out such an analysis in detail.

\(^{22}\) I use 'B-type' and 'C-type' to categorize the violations without suggesting that condition B and particularly condition C as conventionally formulated are the correct account of the violations.
And the same pattern occurs with clefts, where there is a constituent in focus position that is linked to a gap, annotated here as $e$.

(55) a. *It is Mary$_i$ that $e_i$ is proud of her$_i$.
   b. *It is her$_i$, that Mary$_i$ is proud of $e_i$.

(56) a. *It is she/ her$_i$, that $e_i$ thinks we are going to hire Mary$_i$.
   b. *It is Mary$_i$, that she$_i$, thinks that we are going to hire $e_i$.

The generalization therefore appears to be the following: a ‘displaced’ constituent fills the functions of the gap with which it is linked, including its behavior with respect to the binding conditions.

Consider next the following. The a-example is a condition C-type violation, while the b-example is a paraphrase with the same logical form that is not a condition C-type violation.

(57) a. *He$_i$ should do [what(ever) John$_i$ wants to do].
   b. [What(ever) John$_i$ wants to do], he$_i$ should do $t_i$.

In this case, what we have done is topicalize whatever John$_i$ wants to do. Crucially, there is no reconstruction effect.\textsuperscript{23} The linear ordering of the antecedent before the pronoun permits the dependency. What is particularly striking about this case is that when the order is pronoun first, and then antecedent, and everything else is the same, that is, the syntactic structure, the wh-chains and the interpretation, as in (57a), the condition C-type violation occurs.

The usual approach to account for reconstruction is to analyze the constituent in A’ position, that is, the topic or the focus of the cleft, as forming a chain with the gap. Then the connectivity effects are characterized in terms of configurations involving the pronoun or anaphor and the relevant parts of the chain.

However, there is well-known evidence that suggests fairly strongly that connectivity is not a function of the syntactic configuration, but of the correspondence between the syntax and the interpretation. Consider pseudo-clefts. The very simple example in (58) illustrates.

(58) [What, Mary ate $t_i$] was a bagel.

Although a bagel is not the direct object of ate (what is), the meaning of this sentence includes the eating relation between Mary and a bagel, whereby the semantic representation corresponding to a bagel is substituted for what as the Theme argument of the relation EAT.

For the binding conditions, the focus of the pseudo-cleft bears the same relationship to the arguments of the wh-clause that it would bear in the corresponding simple sentence. This is in fact what we have just observed for clefts. In (59a), for example, the direct object her cannot be locally bound by Mary (this is condition B), and the same relation appears to hold in the pseudo-cleft in (59b).

(59) a. *Mary$_i$ is proud of her$_i$.
   b. *What Mary$_i$ is is proud of her$_i$.

\textsuperscript{23} A reviewer suggests that this example could be a case of left dislocation with an empty topic instead of a trace, as in (i) (cf. Koster, 1978).

(i) Whatever John wants to do, ZERO he should do.

In this case there is no reconstruction of whatever John wants to do and thus no Condition C violation. While this is an analysis that definitely should be considered, a prima facie argument against it consists of the fact that there are many cases where a DP is topicalized, reconstructs and produces a Condition C violation.

(ii) a. *John’s mother$_i$, he$_i$ should take care of $t_i$.
   b. John’s mother$_i$, ZERO$_i$ he$_i$, should take care of $t_i$.

This being said, the possibility of a Condition C violation is clearly sensitive to the syntactic structure, or to some correlate of the syntactic structure – there is a difference between whatever John wants to do and John’s mother.
That is, the pronoun in (59b) acts as though it is locally bound by Mary, even though there is no syntactic binding relation here.\(^{24}\) Similar observations hold for the reflexive – it functions as though it is a co-argument of its antecedent.

\[(60)\]
\[
\begin{array}{l}
\text{a. Mary} \text{i is proud of herself,} \\
\text{b. What Mary} \text{i is is proud of herself,} \\
\text{c. (the one) who Mary} \text{i is is proud of is herself.}
\end{array}
\]

On the face of it, it does not appear plausible that Mary c-commands her/herself in the pseudo-cleft examples (59b) and (60b,c).\(^{25}\) The phenomenon that we see here occurs quite generally. The following examples are due to Heycock and Kroch (2002).

\[(61)\]
\[
\begin{array}{l}
\text{a. He}_i \text{ had always claimed that John} \text{n was innocent.} \\
\text{b. What he}_i \text{ had always claimed was that John} \text{n was innocent.} \\
\text{c. That John} \text{n was innocent was what he}_i \text{ had always claimed.}
\end{array}
\]

\[(62)\]
\[
\begin{array}{l}
\text{a. He}_i \text{ really missed John} \text{n s dog.} \\
\text{b. What he}_i \text{ really missed was John} \text{n s dog.} \\
\text{c. John} \text{n s dog was what he}_i \text{ really missed.}
\end{array}
\]

The meaning is the same for all three sentences of each group. The only difference appears to be that in (61b) the pronoun precedes the antecedent and allows for the reconstruction effect, while in (61c) the pronoun follows and the reconstruction effect is suppressed. Thus we have a condition C-type violation only when the pronoun precedes the antecedent. Similarly for (62), where the ordering of the pronoun before the antecedent produces a condition B-type violation that disappears when the order is reversed.

Evidence that what is at issue here is really the linear order of the antecedent and the pronoun, and not syntactic reconstruction, is that there is a semantic connection between the constituent containing the antecedent and a gap. But there is no possibility of forming a syntactic chain, such that there is a syntactic relationship between the pronoun and the chain containing the antecedent. That is, in an example like (62), for example, he may c-command the chain containing what, in virtue of c-commanding the trace of what, but he does not c-command John’s dog, which is outside of the S that immediately contains he.

It is important to stress that in these cases, there is no c-command relation holding at syntax that can explain the binding theory violations. But the conceptual structures are the same as in those examples where there is such a c-command relation. This suggests that the binding conditions in fact hold at conceptual structure, and that moreover linear order imposes an additional condition on binding, when it is licensed by the conceptual structure.

4. Processing linear order

The question now is, why should this be? What does linear order have to do with the computation of grammatical dependencies? As suggested at the outset, it is not sufficient to observe that linear order is relevant to acceptability, and to be satisfied with an account along the lines of the Leftness Condition. The answer, I suggest, is that these dependencies are computed in real time on the basis of the discourse structure. The discourse structure, in turn, is dynamically constructed as each word in the string is encountered, which gives rise to the appearance, captured in Chomsky’s

\(^{24}\) An example that is surprisingly acceptable on this view is

\[(i)\]
\[
\text{(the one) who Mary, is proud of is her,}
\]

Her in this example is a focus. In general, we find that contextual factors can improve examples where the pronoun is coreferential with a co-argument antecedent. E.g.,

\[(ii)\]
\[
\begin{array}{l}
\text{A: Do you know Mary? Is there anyone who is proud of her?} \\
\text{B: MARY is proud of her!} \\
\text{B: Only MARY is proud of her.}
\end{array}
\]

\[(iii)\]
\[
\text{Everyone is proud of Mary – Even SHE is proud of Mary!}
\]

It remains to be seen how to incorporate the interpretation of focus and contrast into the account developed here.

\(^{25}\) There is of course a workaround, which is to assume that the focus is the visible part of a complete sentence that contains the antecedent; an early proposal along these lines is due to Ross (1972). See also Den Dikken (2005).
Leftness Condition, that left-to-right order plays a role in the grammar per se. In some cases, such as WCO, I suggest that linear order is sufficient to account for the acceptability judgments. In other cases I suggest that the processing of linear order interacts with (conceptual) structure conditions on binding.

While it is beyond the scope of this article to try to fully work out an account of how this processing proceeds, it is possible to sketch the outlines of how a discourse representation might be constructed and how referential dependencies are established in terms of this discourse representation. Following Heim (1982), I assume that a noun phrase triggers the creation of a 'file' in the discourse representation. This representation is more or less informative, depending on factors such as those bearing on 'specificity' discussed earlier. The more information a discourse referent has associated with it, the more accessible it is for the formation of referential dependencies. Subsequent processing of the sentence may supply additional information about the noun phrase. I assume that the informativity of an entry in the discourse representation determines its accessibility as an antecedent of a dependent at the point at which the dependency is being processed (see references discussed below). I assume, further, that the acceptability of the dependency is a function, at least in part, of the accessibility of the antecedent.

Precisely why the informativity of the discourse reference should make it more accessible is a question that can only be answered in the framework of a well-understood architecture for discourse representation and retrieval, which I do not have. But there is substantial experimental research that shows that discourse referents vary in terms of their accessibility to retrieval. This experimental research is relevant to the current proposal, because it lends substance to the idea that the acceptability of binding relations in sentences is in part a function of the accessibility of the antecedent.

Among the variables that experimental results show play a role in accessibility are the content of the discourse referent, the syntactic structure in which the corresponding noun phrase appears, and the properties of the probe noun phrase. For instance, Filik and Sanford (2008) studied what happens at the point when the pronoun is encountered and there is no antecedent in the preceding discourse or the context. They found, not surprisingly, that the processor experiences difficulty in such situations. It appears that when the processor fails to find an immediate antecedent, it establishes a discourse referent for the pronoun and then seeks to fix its reference as processing of the sentence proceeds. Moreover, when the processor looks for an antecedent that follows the pronoun, it excludes from consideration those potential antecedents that are impossible for grammatical reasons (Kazanina et al., 2007). These results show that linear order is involved in the computation of referential dependency.

Ariel (1988, 1991, 1994, 2001, 2004) argues that the form of the anaphoric expression used in discourse is sensitive to the accessibility of the antecedent, measured in terms of how much information the processor has about the antecedent at the point at which the anaphoric expression is encountered. Almor (2000:345) argues that the form of the anaphoric expression reflects the processing cost of matching it to the antecedent: "the less specific the representation of the anaphor with respect to the representation of the antecedent, the less costly the anaphor is to process". On this view, a pronoun will be preferred when the antecedent is immediately accessible. Correspondingly, the processing complexity of a pronoun is greater when its antecedent is not accessible at the point at which the pronoun is introduced, a view consistent with the work noted in (i) above. These results show that accessibility is involved in the computation of referential dependency.

Similar notions appear in the literature on discourse processing; see Gernsbcher and Foertsch (1999), Sanders and Gernsbcher (2004), Garnham (1981, 1999), Garnham and Oakhill (1996, 1997), Gundel et al. (1993) among many others. McKoon et al. (1993) provide evidence that the linking of a pronoun to an antecedent can be characterized in terms of a discourse model in which the accessibility of entities in the model varies on the basis of explicit content, contextual information and pragmatic knowledge (72). Gordon and Hendrick (1998) argue that the accessibility of antecedents can be formalized in terms of the structure of the discourse within the framework of Discourse Representation Theory. Finally, Gordon and Hendrick (1997) showed that, the classical Binding Conditions notwithstanding, native subjects show a strong left-to-right bias in the processing of referential dependencies in discourse. Crucially, when a pronoun precedes an antecedent and is not in a subordinate clause (as in Her sister/She visited Lisa at college), coreference is significantly unaccepteable, regardless of whether the pronoun c-commands the antecedent.

---

26 For various approaches to discourse representation and its construction, see among many others Heim (1982), Kamp (1988), Hess (1991), Muskens (1996), Smith (1989), Asher and Lascarides (2003), Asher (1986), and Bos (2003). For related ideas, especially regarding the interactions between syntactic structure, prosody and discourse representations, see Göbbel (this volume), Rochemont (this volume) and Ertschik-Shir et al. (this volume).

27 An open question is the relationship between conceptual structure and discourse representation. As far as I know this question has not yet been raised in the literature. One possibility is that CS representations are constructed in the course of processing and discourse representations are constructed out of CS representations; another is that CS representations are a component of discourse representations. In the absence of a useful foundation to build on, and given limitations of space, I refrain here from further speculation.

28 Varying the syntactic structure results in differences in computing the dependency. This does not mean, however, that the dependency is properly formulated in terms of the syntactic structure, or whether it is a consequence of some aspect of representation that is itself dependent on the syntactic structure.
The experimental results thus give reason to believe that an account of the linguistic data discussed in this chapter can be given an explanation in the form of an explicit discourse processing model.

At the same time, the data in section 2.1 suggest that whether or not a potential antecedent has a thematic role assigned to it also plays a role in the accessibility of the corresponding discourse referent. This fact suggests a possible parallelism between the phenomenon observed here and those addressed in Centering theory (Grosz et al., 1983, 1995). Centering theory is concerned with the relative accessibility of potential antecedents across sentences. Arguments appear on an ordered list determined by their grammatical function. As summarized by Beaver (2003:7), “the subject is the least oblique argument, and becomes the first element of the forward-looking center list. By virtue of this privileged position, it is also termed the preferred center, C₀. The remainder of the forward-looking center list consists of the direct object, then indirect objects, and then adjuncts.” If we assume that the resolution of pronominal dependency within a sentence is subject to the same or similar constraints as the resolution of pronominal dependency across sentences, ceteris paribus, then it should come as no surprise that the least accessible potential antecedent is one that has no grammatical function whatsoever at the point at which the pronoun is processed, namely a wh-phrase in A' position.

Applying these ideas to the cases that we have considered, that is, WCO, binding and reconstruction, we see that the same factors are at play in all of them. In all cases, constraints such as the binding conditions that apply at the point at which the pronoun is processed constrain its dependence on potential antecedents that are available at that point. If the potential antecedent is relatively inaccessible, as in WCO, or in cases where it follows the pronoun, referential dependency is difficult to establish and acceptability judgments are negative.

With this in mind, consider the classic examples illustrating WCO, as in (1), repeated here.

(1) a. “Who_i did his, mother offend t_j?
   b. Who_i offended his, mother?

In (1a) the pronoun is encountered at a point at which there is only a file established for who_i, but no information associated with it that would make it accessible. In contrast, in (1b), at the point at which the pronoun is encountered, who_i has the thematic role associated with the subject of offend and (by hypothesis) is accessible to the pronoun for the purpose of binding.

In case of constructions in which a reflexive can appear initially in the clause, the first potential antecedent is taken to be the actual antecedent, and unacceptability arises when there is lack of agreement. Consider (51), repeated here.

(51) These pictures of \{himself, herself\} are what_i would be tough(est) for John_i to file t_j without Mary_k first looking at pg_j.

The reflexives himself/herself are encountered first in the course of processing, and trigger the construction of a discourse representation in which there is a picture of a male or female individual, respectively. The first potential antecedent is John, which agrees with himself but not with Mary; thus herself produces a mismatch. But where there is no topologicalization, the potential male and female antecedents are both available at the point at which the reflexive is encountered.

(63) What_i would be tough(est) for John_i to file t_j without Mary_k first looking at pg_j are these pictures of \{himself, herself\}.

Similar observations hold for other cases in which the binding conditions are relevant. In pseudo-clefts, for example if the potential antecedent precedes the pronoun, dependency can be established, but if the pronoun precedes the potential antecedent, the lack of a suitable antecedent within the clause that contains the pronoun forces its dependency to be extrasentential. The examples in (61) illustrate the point.

(61) a. He_{i,j} had always claimed that John_i was innocent.
   b. What he_{i,j} had always claimed was that John_i was innocent.
   c. That John_i was innocent was what he_{i,j} had always claimed.

In (61c) John is established as a discourse referent at the point at which he is encountered. At this point, the structure has not been built in which he_{i,j} c-commands the trace of what, so what is not yet linked to that John was innocent. As syntacticians we can see that there should be a condition C violation, but the processor does not have enough information at this point to draw this conclusion, and so John can be the antecedent of he. Crucially, the discovery after the pronoun is processed that (a) the binding conditions are violated or (b) there is a new potential antecedent appears to play no role – dependency is apparently not recomputed.
Consider, finally, examples (17a)–(18a), repeated here.

(17) a. ??The person who, his, mother loves $i$ most is no one.
   (18) a. The person who, his, mother loves $i$ most is John.

In example (17a), the person (who) does not produce a sufficiently accessible antecedent for his, which is thus interpreted as referring to some other discourse antecedent. The quantified expression cannot supply the required referential information. However, in (18a) the identity of the discourse reference can be supplied by the referential John. Note that when the processor can see that there is a condition C violation, as in (61a,b), the subsequent availability of a potential antecedent does not override the constraint.

It thus appears that, for these cases at least, if the binding conditions are violated at a particular point in the processing where the pronoun has already been encountered, the interpretation is ruled out. But if a particular interpretation does not violate the binding conditions at the point at which it is constructed in the course of processing, it is acceptable. Whether a more general conclusion is possible is a question that must remain open at this point.

5. Conclusion

I have brought together a number of phenomena that suggest that the temporal order of processing constituents of a sentence contributes to whether or not a given expression can serve as the antecedent of another, dependent expression. I have focused on the empirical evidence that suggests that it is linear order, coupled with conceptual structure, that determines the acceptability of a range of cases involving the referential dependency of pronouns, anaphors and interrogatives. Many of the phenomena that I have addressed here have been accounted for in strictly syntactic terms in linguistic theory, e.g., weak crossover and superiority. Others, such as Shan and Barker (2006), have suggested that it is possible to understand related phenomena in dynamical terms, that is, as the consequence of the processing of dependency relationships in time as constituents of a sentence are constructed. I have presented additional evidence that suggests that this is a useful way to understand what have previously been viewed as strictly grammatical phenomena, perhaps reflecting grammatical universals to the extent that there is no apparent empirical basis for learning them.

There is of course much more to be done. My proposal is a very preliminary one, based primarily on considerations of syntactic theory and using my own judgments and those from the literature. A broader experimental study should be carried out to validate and refine the acceptability judgments. Beyond that, a real-time experimental study of the processing of pronominal dependency is critical to establishing that in fact there are temporal factors at play. And finally, and most importantly, it is essential to construct a model of the discourse that incorporates an account of how it is constructed and how antecedents are retrieved and dependency computed in order to explain the course of processing and the judgments that follow.

If the program sketched out in this paper is on the right track, the consequence is that it facilitates two major steps forward in our understanding of how language works. First, it allows for a dramatic simplification of syntactic theory, by removing from the domain of syntax a number of principles that are specific to syntactic configuration. Moreover, it removes the motivation for particular complexities of syntactic structure that have been motivated in large part by the fact that they permit these configuration-specific principles to apply. This step is fully aligned with the program of Simpler Syntax, Culicover and Jackendoff (2005).

Second, it offers the possibility of a deeper explanation for a fairly intricate set of facts that at best could be organized and codified under a set of configuration-specific principles. The perspective that the current proposals have been contrasted with was firmly established in the Conditions framework of Chomsky (1973) and has been the dominant view of such phenomena as weak crossover in syntactic theory during the last century. These principles establish important and far-reaching generalizations; however, they leave open the question as to why these generalizations might hold. What I have proposed here by no means accounts for all of the phenomena that have been addressed during that period. But it

---

29 I have not considered, for example, the phenomenon of ‘anti-reconstruction’, which appears to show the role of both linear order and structure. E.g.,

(i) a. *[Which pictures of John?] does he, like $i$, the best?
   b. *[Which pictures that John, painted] does he, like $i$, the best?

On reconstruction, both (i.a) and (i.b) should be ruled out by condition C. But if linear order rules in (i.b), then it should also rule in (i.a). However, the structural relationship of John to the fronted NP is different in the two cases. It is possible that the difference is reflected in a difference in accessibility at the level of discourse representation, but I have no specific proposal along these lines that would be anything more than a recoding of the syntactic difference. It should also be noted that not all speakers get a difference between (i.a) and (i.b).
does appear that it is possible to replace certain generalizations formulated in configurational terms with explanations for why they hold, or appear to hold, in terms of independently required mechanisms of processing sentences and building representations of their meaning in time. To the extent that this effort is successful, it provides additional reason to be optimistic that the programs of Kluender and Hawkins to account for apparent grammaticality constraints in terms of processing complexity, and that of Erteschik-Shir to account for apparent constraints in terms of discourse structure, may ultimately prove to be correct.

Acknowledgments

I am extremely grateful to Susanne Winkler, Tom Wasow, Ivan Sag, Jutta Hartmann and several anonymous reviewers for their close reading of the manuscript and for numerous pertinent comments, suggestions and advice that have made this article much better than it started out to be. The responsibility for any remaining problems is mine, of course.

References


