

The Qualitative Difference in Acceptability in Extraposition from NP Sentences*

Kayono Shiobara

University of Tokyo / University of British Columbia

kayono@interchange.ubc.ca

Abstract

Extraposition from NP (EXNP) sentences in English exhibit a variety of syntactic, semantic and pragmatic properties. Shiobara(1997) analyzes EXNP in terms of sentence processing and accounts for the acceptability of EXNP sentences by syntactic and semantic/pragmatic processing principles. However, the degree of acceptability of each sentence is not necessarily the same. For example, an EXNP sentence with preposition-stranding is always found to be worse than a corresponding EXNP sentence with pied-piping of a preposition. This paper shows that the qualitative difference in acceptability in EXNP sentences results from the modified version of the Nested Dependency Constraint proposed by Fodor(1978).

Keywords: Extraposition from NP (EXNP), preposition-stranding, pied-piping of a preposition, processing, the Nested Dependency Constraint

1. Introduction

In Extraposition from NP (EXNP) sentences, pied-piping of a preposition (p-piping) is preferred over preposition-stranding (p-stranding).

- (1) (Did a book about the life of Diana come out ?)
 - a. No. But a book about that accident in Paris came out yesterday.
 - b. No. (?)But a book came out yesterday about that accident in Paris.
 - c. No. **But a book about came out yesterday that accident in Paris.
- (2) (Were you talking about the wife of Bill when I saw you on the street ?)
 - a. No. We were talking about the wife of John at that time.
 - b. No. We were talking about the wife at that time of John.
 - c. No. (?)We were talking about the wife of at that time John (, and at another time, we were talking about Bill's wife).

The (b) and (c) examples in (1) are extraposition from subject and those in (2) are extraposition from object. The extraposition from subject with p-stranding in (1c) is absolutely unacceptable whereas the extraposition from object with p-stranding in (2c) is not so bad especially when the adverbial phrase *at that time* is interpreted parenthetically.¹ Note that it is not the case that what the sentence in (1c) shows is that extraposing an NP is prohibited because the following sentence is fully acceptable.

- (3) I pressed the trigger and a hole appeared in his forehead [_{NP} the size of a quarter].

(Bache and Nielsen(1997:156))

In (3) the sentence-final NP *the size of a quarter* modifies the noun *hole*. Notice also that the corresponding wh-movement with p-stranding is awkward but not as bad as (1c).

- (4) ??What did a book about – come out yesterday ?

Thus, the unacceptability in (1c) shows that (i) extraposition of an NP (ii) from subject (iii) with p-stranding is prohibited. The factors in (i)-(iii) altogether make a sentence seriously unacceptable.

Shiobara(1997) attempts to account for the contrast between p-piping and p-stranding in terms of syntactic processing principles proposed by Hawkins(1990,1993): namely, the principle of Early Immediate Constituents (EIC). The EIC correctly predicts that the examples in (c) are harder to process than the corresponding (b) examples as we will see in section 2. However, it is incomplete in that it does not explain the qualitative

* I am grateful to Noriko Imanishi, Randy Sharp, Hidekazu Tanaka, Chris Tancredi and Akira Watanabe for valuable comments and suggestions. I would like to thank as well the audience at the annual meeting of University of Tokyo English Linguistics Association in July 1999 where parts of this paper was presented.

¹ It is generally agreed that extraposition from subject is more restricted than extraposition from object. Let us look at the close minimal pairs in (i) and the sentence in (ii).

- (i) a. *A man hit Mary yesterday with green eyes.
- b. John hit a girl yesterday with green eyes.
- (ii) The bookseller hit the boy yesterday with a book in his hand.
- Interpretation :
- a. *The bookseller with a book in his hand hit/punched the boy.
- b. The bookseller hit the boy who had a (stolen) book.
- c. The bookseller used a book to hit the boy.

In (i) the extraposed PP with green eyes can not be associated with the subject as in (ia) but can be associated with the object as in (ib). In (ii) the extraposed PP can be associated with the object or interpreted as a VP-modifier, while subject-host reading is not obtained.

difference in acceptability between (1c) and (2c) on the one hand (i.e. factor (ii) above), and between (1b) and (1c) on the other (i.e. factor (iii) above), that is, the fact that the sentence in (1c) is found much worse than (2c) or (1b). This paper sheds light on these contrasts and proposes a processing explanation for them.

Section 2 takes a look at the processing principles proposed by Hawkins and how they apply to EXNP sentences. In section 3, by generalizing and modifying the Nested Dependency Constraint proposed by Fodor(1978), I account for the preference of p-piping to p-stranding and show that the same constraint can explain the difference in acceptability between extraposition from subject and extraposition from object. Section 4 is the conclusion.

2. EXNP and Syntactic Sentence Processing Principles

2.1. *The Processing Mechanism in Hawkins(1990,1993,1994)*

Syntactic processing concerns how the sequentially perceived words are grouped into constituents hierarchically. This grouping (i.e. parsing) is done with quite remarkable speed and efficiency. Hawkins explains this rapidity and efficiency of on-line syntactic processing by positing two fundamental considerations. First, he notices that there are some syntactic categories that are unique to their respective dominating mother nodes and that can act as signals to the parser to construct the relevant mother node. For example, such function words as Det(erminer) and Comp(lementizer) construct NP and S' respectively, and heads uniquely construct phrases (e.g. N is unique to NP, V to VP etc.). He accordingly proposes the following principles of constituency recognition:

(5) Mother Node Construction (MNC)

In the left-to-right parsing of a sentence, if any syntactic category uniquely determines a mother node M (or nodes) in accordance with the phrase structure rules of the language, then this M is constructed over the category in question, immediately and obligatorily.

(6) Immediate Constituent Attachment (ICA)

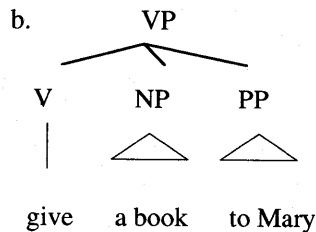
An immediate category IC that can be attached to an M in accordance with the phrase structure rule of the grammar will be so attached, as rapidly as possible. ICs may be encountered after the category that constructs the M in the left-to-right parse, or they may be encountered before it [...]. In either event the attachment of all sister ICs to an M, either backward or forward,

proceeds as fast as possible.

(Hawkins(1990:228))

For example, in English, when we encounter the verb *give*, a mother node VP is constructed immediately and obligatorily by the MNC in (5), and then the complement NP and PP are attached to the mother node VP as rapidly as possible by the ICA in (6):

(7) a.... give [NP a book] [PP to Mary]



The MNC assumes the following two axioms: the Axiom of Existence, which says that for each phrasal mother node M, there exists at least one daughter category C that can construct M on each occasion of use; and the Axiom of Uniqueness, which says that each mother node constructing category (MNCC) will consistently construct a unique M on each occasion of use. (cf. Hawkins (1993:240-241))

Second, and the main argument in Hawkins(1990), is that the rapidity in parsing depends on the order of words. He proposes the following principle:

(8) Early Immediate Constituents (EIC)

The human parser prefers to maximize the left-to-right IC-to-word ratios of the phrasal nodes that it constructs.

(ibid:233)

The term “left-to-right IC-to-word ratio” is defined as in (9a) and the term “constituent recognition domain” in (9a) is defined as in (9b):

(9) a. Left-to-Right IC-to-Word Ratio

The left-to-right IC-to-word ratio for a constituent recognition domain is measured by first counting the ICs in the domain from left to right (starting from 1) and then counting the words in the domain from left to right (again starting from 1). For each word and its dominating IC, the IC total is divided by the word total at that point, and the result is expressed as a percentage. ... The higher the percentage, the more

loaded and informative is the constituency information at that point. An aggregate IC-to-word ratio for the whole constituent recognition domain is then calculated by averaging the percentages for all the words in the domain. The higher the aggregate, the more optimal is that order of words for processing. The ratio for a whole sentence can be defined as the average of the aggregate IC-to-word ratios for all constituent recognition domains in the sentence, that is, for all phrasal categories that the sentence dominates.

(ibid:233-234)

b. Constituent Recognition Domain (CRD)

The constituent recognition domain for a node X is the ordered set of words in a parse string that must be parsed in order to recognize all ICs of X, proceeding from the word that constructs the first IC on the left, to the word that constructs the last IC on the right, and including all intervening words.

(ibid:229)

The EIC in (8) is based on the observation that the heavier (in length, regardless of content) an element is in English, the more frequently it will be postposed.² The main claim in Hawkins(1990) is that the EIC is one of the innate processing mechanisms, which is reflected in cross-linguistic regularities of word order. As for reordering rules including EXNP, EIC makes the following prediction:

(10) EIC Potential Increase Prediction

If a grammar encodes a rearrangement rule in response to EIC (for example, Heavy NP Shift), the rule will apply only to those grammatical categories and only in that (leftward or rightward) direction that will potentially increase the EIC ratios of the relevant CRDs. ...

(ibid:246)

The concept of the CRD in (9b) is crucial to syntactic processing mechanisms. To put it simply, the CRD corresponds to various dependencies in a sentence that can not be recognized in the mere sequence of words. For example, the CRD for main clause S

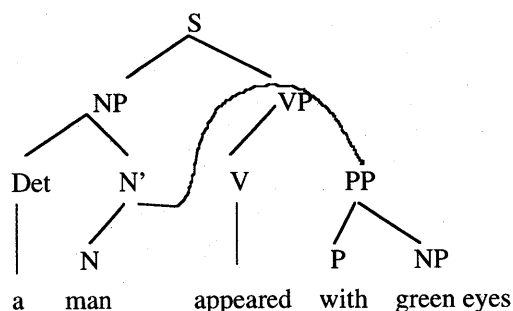
² Hawkins(1994) alters the way of calculating ratios. Details aside, under the new definition, a CRD consists of the set of non-terminal nodes as well as terminal nodes, i.e. words, and the calculation itself becomes simpler. I intentionally adopt the version in Hawkins(1990) because I do not assume the exact same structure as Hawkins(1994).

consists of at least a subject NP and a verb, and it signifies the dependency of predication.

Shiobara's(1997) claim is that these independently motivated principles, especially the EIC, predict not only the acceptability of EXNP sentences, but also the reordering possibilities of extraposed phrases. In order to apply the EIC to the dependency between an extraposed phrase (i.e. modifier) and its host noun (i.e. modifiee), we are forced to assume the axiom of phrase structure which allows branches to cross, proposed in McCawley(1982):

- (11) a. A man appeared with green eyes.

b.



Hawkins(1994, 1998:749) argues in favor of this kind of phrase structure by presenting a piece of evidence that concerns the binding possibility of nouns. Judging from his argument for the structure, he must assume that this phrase structure is allowed in the grammar. I assume here that this kind of structure is composed on-line during sentence processing by a hearer and is a processing-particular structure. For example, when we encounter a preposition of an extraposed PP during sentence processing, we immediately begin to look for its host noun and establish a dependency between them in the fashion illustrated in (11b). I would like to leave it open whether this kind of structure is allowed in the grammar or not.³

³ As for the syntactic structure of an extraposed phrase, syntactic tests such as VP-deletion and VP-fronting suggest that a phrase extraposed from subject attaches to S while a phrase extraposed from object attaches to VP. (cf. Guéron(1980)) Another possibility is that an extraposed phrase is not attached to the sentence at any syntactic levels and gets associated with its host noun semantically. (I thank Masaru Kajita for suggesting this possibility to me.) This view is supported by the fact that there are sentences where both subject-host and object-host readings can be obtained, though object-host reading is preferred (cf. Note 1):

(i) Some girl slapped some boy yesterday [who was...] (Frazier and Clifton(1996:33))

In (i) the extraposed relative clause can be associated with subject or object.

2.2. EXNP in Hawkins' system

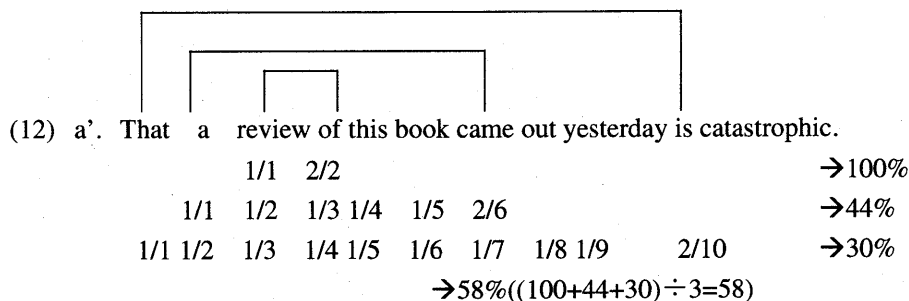
2.2.1. EXNP as a Sentence Processing Strategy

Before looking at how EIC applies to EXNP sentences, let me make it clear that I do not assume EXNP is generated by a transformation. EXNP is regarded as a processing strategy which makes it easy for a hearer to process a sentence by achieving "end-weight" and/or "end-focus" (cf. Quirk et al. (1985:1398)). I take the null hypothesis to be that EXNP is not subject to syntactic constraints at all and it does not involve movement. If correct, EXNP is never a source of ungrammaticality. Rather, what distinguishes acceptable EXNP sentences from unacceptable ones are processing principles. This is the analysis I am pursuing.^{4,5}

2.2.2. EXNP and EIC

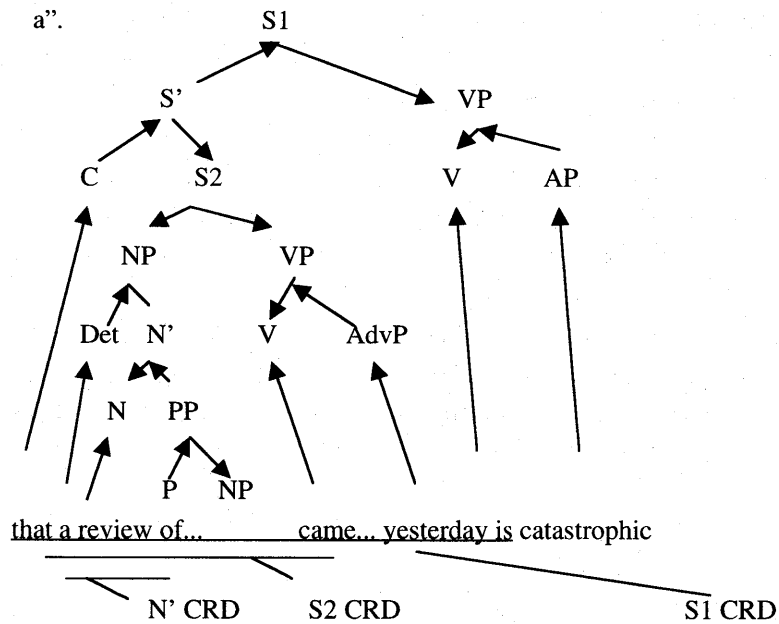
Now we are ready to see how the EIC applies to EXNP examples.

- (12) a. That a review of this book came out yesterday is catastrophic.
 b. (?)That a review came out yesterday of this book is catastrophic.
 c. *That a review came out yesterday is catastrophic of this book.

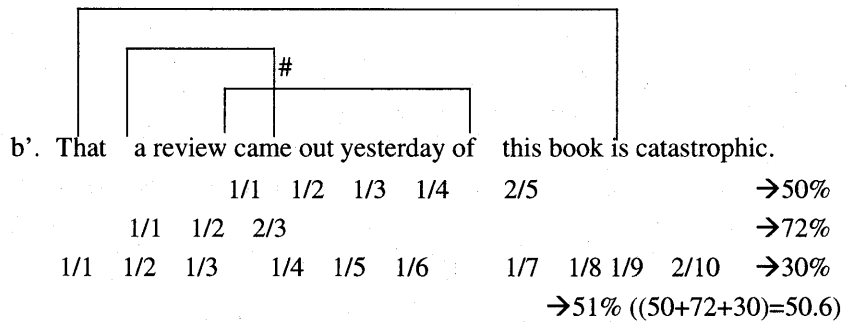


⁴ See Shiobara(1997:chapter4) for the discussion.

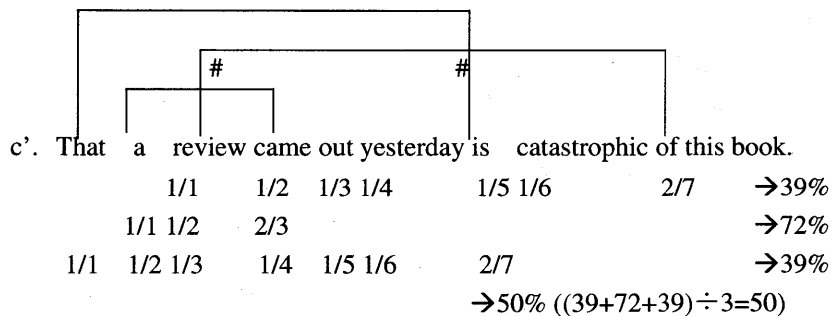
⁵ I continue to use the term "extraposition" for the sake of convenience, though the term is misleading given the assumption that extraposition does not involve movement.



(NB: The upward arrows show how Ms are constructed by ICs or how ICs are attached to M ; the downward arrows show how ICs are constructed by Ms.)



EXTRAPOSITION FROM NP



The sentence in (12a) does not involve EXNP; in (12b) the PP *of this book* is extraposed to the final position of the sentential subject; in (12c) the PP is extraposed to the sentence-final position. The sentence in (12a), for example, is processed in the following way (See (12a'')). The first perceived word *that* indicates that the sentence has a sentential subject S' (unless it is immediately followed by a noun) so it constructs the mother node S', which in turn constructs S1 and S2. Next, the determiner *a* attaches to the mother node NP, which was constructed by the mother node S2. Then the noun *review*, the PP *of this book* and the verb *came* are attached to the relevant nodes and finally the adjective *catastrophic* comes, indicating a predicate adjective, and the AP is attached under the mother node VP.

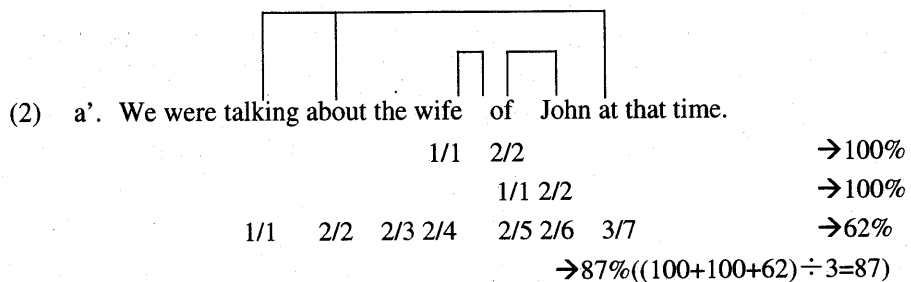
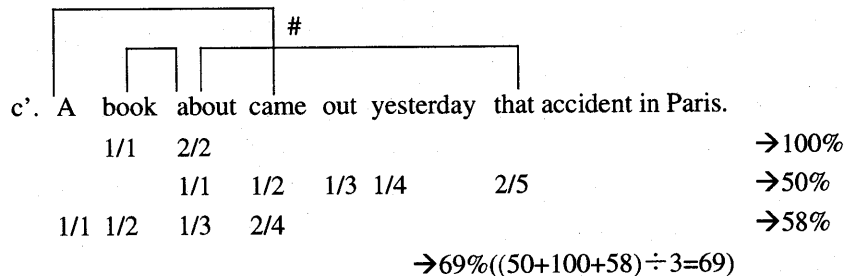
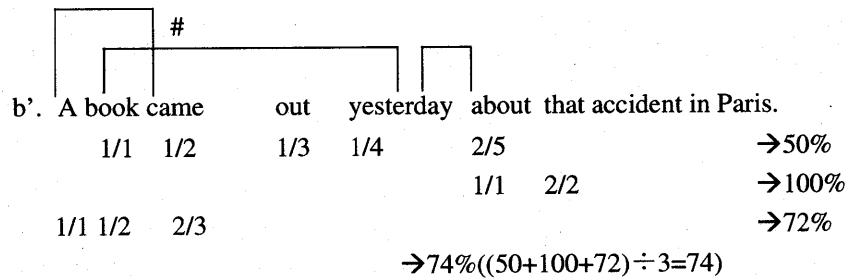
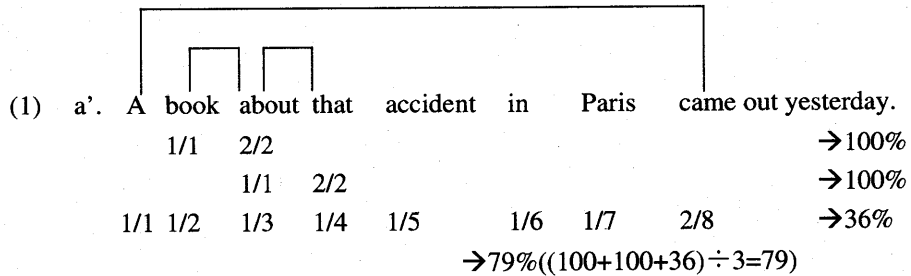
Let us compare (a)-(c) in (12). The relevant Constituent Recognition Domains (CRDs) are the N' consisting of the extraposed PP *of...* and its host noun *review*; the sister ICs of S2 (i.e. the NP *a review...* and the VP *came out...*), and the sister ICs of the entire clause S1 (i.e. the sentential subject *that...* and the VP *is...*). The aggregate ratio of each sentence is 58%, 51% and 50% respectively, which correctly predicts that the EXNP sentence in (12b) is harder to process than the sentence in (12a), and in turn, the EXNP sentence in (12c) is harder to process than (12b), though the aggregate ratio does not show a drastic difference between (b) and (c) (51% vs. 50%).⁶ Note that two CRDs intersect in (12b) and (12c) as the double-cross mark shows, which will be relevant in the next section.

Thus, the EIC is strong in that it makes right predictions concerning not only whether EXNP is possible or not ((a) vs. (b)), but also where an extraposed phrase can appear ((b) vs. (c)). The higher the aggregate ratio, the more optimal is that word order of processing, and the more acceptable, in turn, the sentence is. Note also that the higher

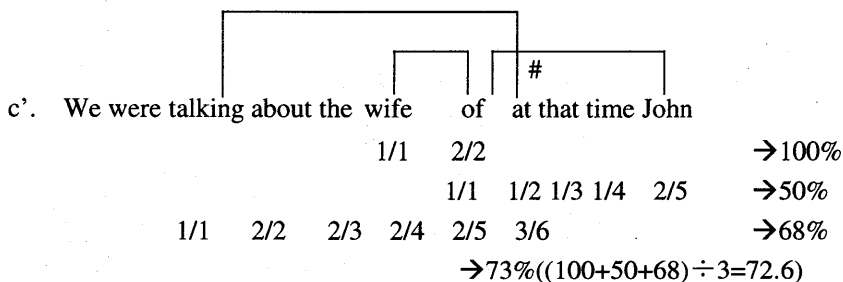
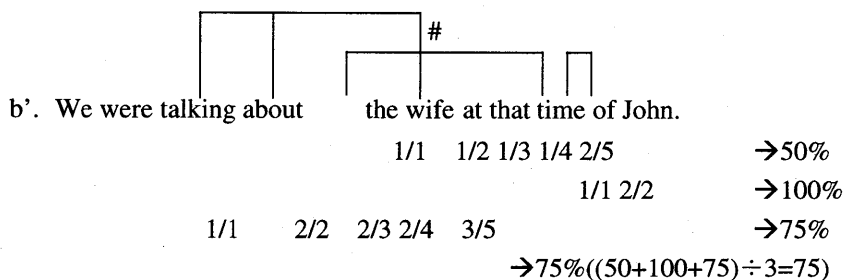
⁶ There are many other CRDs in a sentence, of course, but we have only to calculate the ratios of relevant CRDs in comparing sentences. What is important is not the absolute value of the aggregate ratio of each sentence, but whether the ratio of one sentence is higher than that of the others.

the number of intersecting CRDs (#), the less acceptable the sentence is. I will come back to this point in Section 3.

The aggregate ratios of the examples in (1) and (2) are calculated in the same fashion.



EXTRAPOSITION FROM NP



As for the examples in (1), the EIC rightly predicts that the sentence in (1c) is harder to process than (1b), and (1b) is harder than (1a), because the aggregate ratio is the lowest in (1c) and the highest in (1a). In (2), however, the example in (2b) is wrongly predicted to be difficult to process because the ratio is lower than that of (2a). The decline of ratio in (2b) is due to the length of the adverbial phrase *at that time* which makes the N' CRD in (2b) bigger than that in (2a).

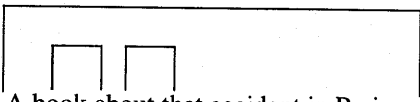
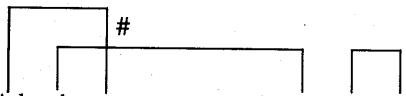
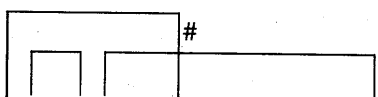
As is obvious from this argument, the weakness of the EIC is that it takes only the size of CRDs (i.e. number of words) into consideration and does not see what kind of CRD it is.⁷ The sentence in (1b) is worse than (1a) intuitively because the predicate *came out yesterday* intervenes between the extraposed phrase *about...* and its host noun *book*, while the sentence in (2b) is all right because it is a parenthetical adverbial phrase that separates the extraposed PP from its host. In the next section, I will formalize this intuitive description by referring to the crossing of CRDs.⁸

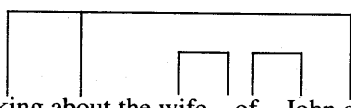
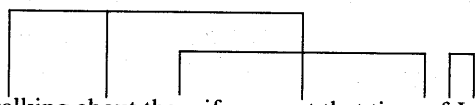
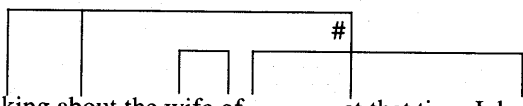
⁷ Another obvious problem with EIC is in its word-counting system. EIC counts every morphological word equally whether it is prosodically heavy or light. Shiobara(1999) proposes a prosodic version of EIC.

⁸ EIC does not make any prediction regarding the contrast among the sentences that contain different words (e.g. (1c) vs. (2c)). I suppose that semantic/pragmatic processing principles work in predicting the different acceptability of such cases. (See Frazier and Clifton(1996) and Shiobara(1997) for a discussion of semantic/pragmatic processing principles.)

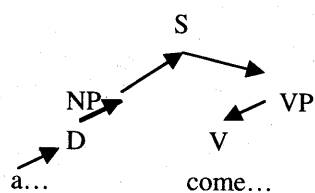
3. Crossing of CRD-dependencies

Let us reexamine the relevant CRDs which the examples in (1) and (2) contain.

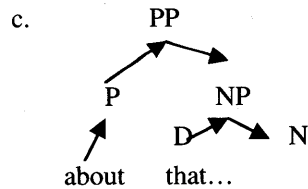
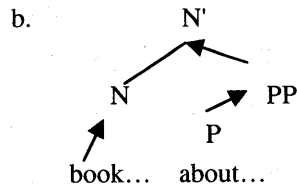
- (1) a.  A book about that accident in Paris came out yesterday.
- b.  (?)A book came out yesterday about that accident in Paris.
- c.  **A book about came out yesterday that accident in Paris.

- (2) a.  We were talking about the wife of John at that time.
- b.  We were talking about the wife at that time of John.
- c.  (?)We were talking about the wife of at that time John.

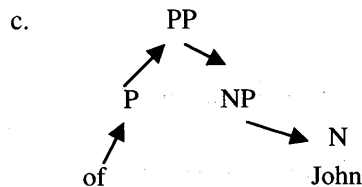
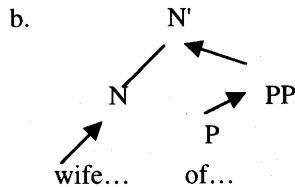
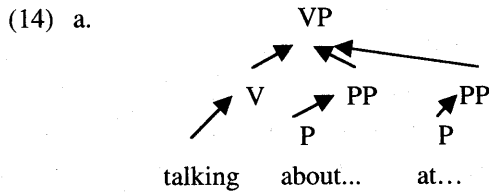
The relevant CRDs in (1) are: (13a) the entire sentence S which consists of subject and predicate, (13b) N' which consists of a noun and a PP, and (13c) PP which consists of a preposition and its complement NP:

- (13) a. 

EXTRAPOSITION FROM NP



The relevant CRDs in (2) are: (14a) the VP which consists of a verb, an optional PP and an adverbial PP, (14b) N' and (14c) PP:



In (13a), the determiner *a* signals a subject NP, which constructs the mother node S, and S requires a predicate VP. In (13b) and (14b), a noun does not necessarily take a PP complement and an N' node is constructed only when there is a complement. In (13c) and (14c), on the other hand, a preposition requires a complement NP and a node N is constructed when the preposition is processed. In (14a), as in (13b) and (14b), the verb

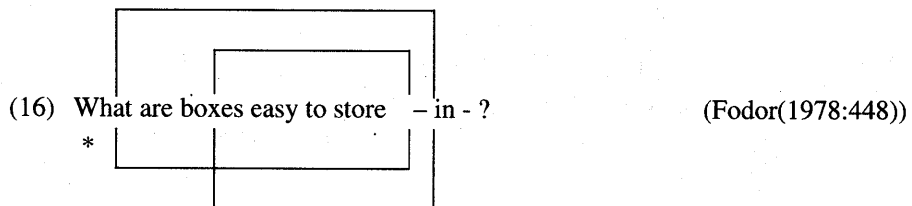
talk does not need a PP complement, and PP nodes are constructed and attached under VP only when a preposition is encountered and processed. Therefore, the CRDs in (13a), (13c) and (14c) are strong in the sense that the first IC of the CRD requires the existence of the second IC, as the downward arrow shows. In contrast, the CRDs in (13b), (14a) and (14b) are weak in that the second (and third) IC is optional. My claim is that this kind of qualitative difference in CRDs, which is ignored in Hawkins' system, must be related to the qualitative difference in acceptability of EXNP sentences. Now, I would like to introduce another principle proposed in Fodor(1978).

(15) 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.

(Fodor(1978:448))

Originally, NDC is formulated as a no-ambiguity constraint; that is, it disambiguates sentences that are potentially ambiguous with respect to how their gaps are to be filled. Fodor(1978) argues that NDC has been grammaticalized in response to pressure from the parser and it covers all interactions of movement (whether leftward or rightward) and deletion rules. To take an example, in the following sentence, the nested reading *store boxes in what* is grammatical while the intersecting reading *store what in boxes* is not:



I modify the NDC in (15) as a processing principle in order to apply it to EXNP cases:

(17) The Modified Nested Dependency Constraint (MNDC)

When there are two or more CRD-dependencies in the same sentence, an intersection of the CRD-dependencies leads to processing difficulty.

Here I regard MNDC as a processing principle like EIC, not a principle of grammar, which means that MNDC works as a measure of processing difficulty in a sentence and

the violation of MNDC does not directly lead to ungrammaticality of the sentence. I refer to the CRD-dependency as a relation between sister ICs in a CRD.

Another proposal I make here is that a crossing of strong CRD-dependencies leads to a severe ill-formedness of the sentence which contains it; a crossing of a strong dependency and a weak dependency makes the sentence marginal; and a crossing of weak dependencies is not a problem. The strong/weak CRD-dependency is defined in the following way:

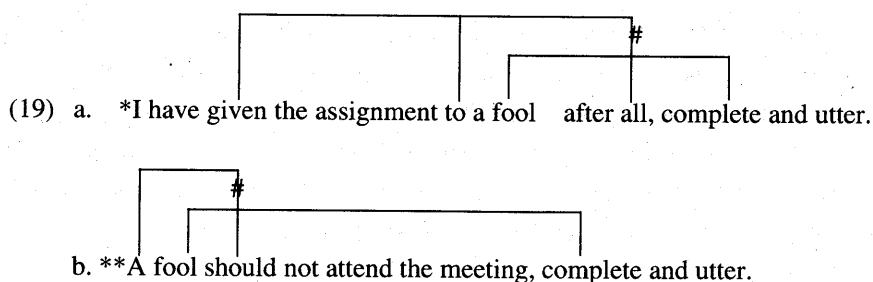
(18) The Strength of a CRD-dependency

A CRD-dependency is strong when the first IC of a CRD requires another IC; weak, in other cases.⁹

Let us go back to examples in (1) and (2) and see how this proposal works together with MNDC. The sentence in (1c) contains a crossing of strong CRD-dependencies, S and PP. The awkwardness in (1b) and (2c) can be reduced to a crossing of strong and weak CRD-dependencies, S and N', PP and VP, respectively. The sentence in (2b) is all right where there is a crossing of weak CRD-dependencies, VP and N'. By referring to MNDC in (17) and the strength of CRD-dependencies defined as in (18), the different acceptability in (1) and (2) can be well described. More generally, it follows that p-piping is preferred to p-stranding because the former forms a weak, and the latter, a strong CRD-dependency. And Extraposition from object is less restricted than extraposition from subject because the former crosses a weak dependency of VP-CRD, and the latter crosses a strong dependency of S-CRD. Therefore, extraposition from subject with p-stranding is predicted to be worse than extraposition from object with p-stranding or extraposition from subject with p-piping, because it always involves a crossing of strong CRD-dependencies, namely a dependency of predication in S and that of PP, whereas the latter two involve a crossing of strong and weak CRD-dependencies.

Let us look at one more example where another type of strong CRD-dependency exists.

⁹ The strength of dependencies might correlate with whether the dependency contains a movement or not if we adopt the VP internal subject hypothesis and take the dependency of predication as in (13a) to involve subject-raising. (I thank Akira Watanabe for suggesting this to me.) That is, a filler-gap dependency is strong while others are weak. This way of capturing strong-weak dependencies implies that the processor is sensitive to the trace of VP internal subject. See Bever and Santz(1997) for the related discussion. I leave this issue for further research.



The AP *complete and utter* is prenominal and it requires a following noun. In this sense, this AP and a noun form a strong CRD-dependency (i.e. N'-CRD). This strong dependency crosses a weak VP-CRD-dependency in (19a) while it crosses another strong S-CRD-dependency in (19b). The sentence in (19b) is as bad as (1c) and the system which I propose here correctly predicts this qualitative similarity of ill-formedness.

4. Concluding Remarks

By looking at the strength of CRD-dependencies in detail and adopting the modified version of NDC in (17), we can see why extraposition from subject with p-stranding is much worse than other EXNP examples. However, properties which EXNP sentences exhibit are not this simple and we need to examine much more data in order to state the strength of dependencies more explicitly. For example, there is a crossing of strong and weak CRD-dependencies both in (2c) and (18a) but the latter is found to be much worse than the former. Our system proposed in Section 3 does not give an explanation for this contrast and needs to be further refined.¹⁰ It might also be worth investigating how the strength of dependency is related to actual processing mechanism and/or the system of working memory.

References

- Bache, Carl and Niels Davidsen-Nielsen (1997) *Mastering English: An Advance Grammar for Non-native and Native Speakers*, Mouton de Gruyter, Berlin.

¹⁰ By introducing MNDC, I do not mean that EIC should be abandoned. Rather, considering the interaction between EIC and MNDC should lead us to a better account of the qualitative difference in acceptability in EXNP sentences.

- Bever, Thomas G. and Montserrat Sanz (1997) "Empty Categories Access Their Antecedents during Comprehension; Unaccusatives in Spanish," *Linguistic Inquiry* 28, 69-91.
- Fodor, Janet (1978) "Parsing Strategies and Constraints on Transformations," *Linguistic Inquiry* 9, 427-473.
- Frazier, Lyn and Charles Clifton (1996) *Construal*, MIT Press, Cambridge, Massachusetts.
- Guéron, Jacqueline (1980) "On the Syntax and Semantics of PP Extraposition," *Linguistic Inquiry* 11, 637-678.
- Hawkins, John A. (1990) "A Parsing Theory of Word Order Universals," *Linguistic Inquiry* 21, 223-261.
- Hawkins, John A. (1993) "Heads, Parsing and Word-order Universals," *Heads in Grammatical Theory* ed. by Greville G. Corbett, Norman M. Fraser and Scott McGlashan, 231-265, Cambridge University Press.
- Hawkins, John A. (1994) *A Performance Theory of Order and Constituency*, Cambridge University Press.
- Hawkins, John A. (1998) "Some Issues in a Performance Theory of Word Order," *Constituent Order in the Language of Europe*, ed. by A. Siewierska, 729-781, Mouton de Gruyter, Berlin.
- McCawley, James (1982) "Parentheticals and Discontinuous Constituent Structure," *Linguistic Inquiry* 13, 91-106.
- Quirk, Randolph, Sydney Greenbaum, Geoffrey Leech and Jan Svartvik (1985) *A Comprehensive Grammar of the English Language*, Longman, London.
- Shiobara, Kayono (1997) *Extraposition from NP as a Sentence Processing Strategy*, MA thesis, the University of Tokyo.
- Shiobara, Kayono (1999) "Prosodic Constraint on Rightward Dislocation and Phonological Phrasing in English," ms. the University of British Columbia.