A Memory-based Sentence Processing Model

Kei TAKAHASHI Kei YOSHIMOTO

Graduate School of International Cultural Studies (GSICS)
Tohoku University

41 Kawauchi, Aoba-ku, Sendai 980-8578, Japan

E-mail: {kei-ta, kei}@linguist.jp

Abstract In the syntactic literature, there are some linguistic data which involve linear order. Such phenomena are difficult for structure-based accounts. LFG succeeds in the prediction of correct grammatical status, however, it fails in accounting for other data since word order is not reflected. Through the sentences in problem, we consider this problem should be solved in terms of real-time processing, not in terms of syntactic manner. Thereupon, we propose a memory-based sentence processing model, which uniformly accounts for the problematic sentences for syntax. Then, we argue the grammar-processor interaction through adopting our model to combitonary categorial grammar.

keywords sentence processing, deactivation, LFG, CCG, topicalization, coordination, RNR, insertion

1 Introduction

As has been made clear, in the linguistic literature, natural language syntax cannot be described *solely* in terms of linear order (finite-state grammar); rather, hierarchical structures play a crucial role. However, that does not mean that linear order has no role to play in our accounting for native speakers' (un)acceptability judgments. To say the least, given that (un)acceptability judgments are results of real-time processing, it would be unnatural to consider that linear order has no effect on such judgments. In fact, some researchers (e.g. Hawkins 1994) have attempted to explain syntactic phenomena in terms of real-time processing.

In this paper we make a a specific proposal about linear order effects on the syntactic phenomena and formulate it in terms of working memory. The structure of the paper is as follows. In Section 2, we illustrate the linguistic data we intend to account for; we point out that the existing syntactic accounts fail to capture the observational generalization behind the constructions. In Section 3, we propose a linear order account, which is modeled in terms of working memory. In Section 4, we demonstrate that our proposal uniformly accounts for the phenomena in question. In Section 5, we attempt to propose the grammar-processor interface. Here, we show how our idea can be enbodied as a parser using combinatory categorial grammar (CCG, Steedman 2000). In Section 6, we discuss the remaining problem and concludes the paper.

2 Data and Problems

Kaplan and Bresnan (1982) pointed out the contrast in (1a,b), which is a problem (at least) for a movement-based analysis of topicalization. An obvious explanation for the ungrammaticality of (1a) would be that *of* cannot take a *that*-clause as its complement. In movement-based analysis, it is expected that topicalization of the *that*-clause does not alter the sentence's (un)acceptable status. However, this expectation is betrayed by (1b).

(1) a.*John was thinking of that he was stupid.

b. That he was stupid, John was thinking of.

The solution proposed in the LFG literature (Kaplan and Zaenen 1989; Bresnan 2000; Falk 2001) is based on the LFG assumption that complement selection is stated in terms of grammatical function (GF), instead of part of speech (POS); the relation between GF and POS is stated by a separate constraint or theory (Lexical Mapping Theory; LMT). The standard account of (1b) in LFG is as follows: (i) TOP (=topic) is mapped to CP [1]; while TOP can be realized as a CP, OBJ (= object) cannot. On the other hand, OBJ can only be realized as an NP by LMT; (ii) *of* selects OBJ; and (iii) the syntactic relation between *of* and the topicalized clause in (1b) is stated in terms of functional uncertainty, resulting in co-sharing the value of TOP and OBJ.

(1a) is bad because, while a *that*-clause attempts to realize OBJ in violation of (i), the GF is selected by *of*. Since the value of OBJ is empty, the "completeness condition" is violated, which constrains that each value of f-structure must not be empty. In contrast, in (1b), the OBJ in question is not realized by an overt expression (i.e., no violation of (i)); indeed, the value of OBJ is the one of TOP, *that*-clause. However, since an f-structure contains no POS information, this does not violate (i).[2]

$$*\begin{bmatrix} \textsc{pred} & \text{`think of}\langle(\uparrow \textsc{subj})(\uparrow \textsc{obl}_{of}\textsc{obj})\rangle\text{'} \\ \textsc{subj} & \begin{bmatrix}\textsc{pred} & \text{`John'}\end{bmatrix} \\ \textsc{tense} & \textsc{past} \\ \textsc{obl}_{of} & \begin{bmatrix}\textsc{pcase} & \textsc{obl}_{of} \\ \textsc{obj} & \underline{\ ?\ }\end{bmatrix}\end{bmatrix}$$

Figure 1: f-structure of (1a)

However, standard LFG accounts fail to predict the grammaticality contrast in the following coordination examples. ((4b-c) are taken from Quirk et al. (1995, §10.41))

(2) a. John was thinking of [his girlfriend].

b.*John was thinking of [that he was stupid]. (=1a)

c. John was thinking of [his girlfriend] and [that he was stupid].

d.*John was thinking of [that he was stupid] and [his girlfriend].

(3) a. Ken agreed with, but John denied, that Mike was wrong.

b.*John denied, but Ken agreed, with that Mike was wrong.

(4) a. Either she or you are/*is wrong.

b. Either your brakes or your eyesight is/?are at fault.

c. Either your eyesight or your brakes are/?is at fault.

(2c-d) differ only in the order of the conjuncts (which are bracketed). In standard LFG, the f-structures of these sentences are one and the same structures, as shown in figure 2.

$$\begin{bmatrix} \textsc{pred} & \text{`think of}\langle(\uparrow\textsc{subj})(\uparrow\textsc{obl}_{of}\textsc{obj})\rangle \\ \textsc{subj} & \begin{bmatrix}\textsc{pred} & \text{`John'}\end{bmatrix} \\ \textsc{tense} & \textsc{past} \\ \textsc{obl}_{of} & \begin{bmatrix}\textsc{pcase}\ \textsc{obl}_{of} & \\ \textsc{obj} & \left\{\begin{matrix}\textsc{pred} & [\text{`his girlfriend'}] \\ \textsc{pred} & [\text{`that ...'}]\end{matrix}\right\}\end{bmatrix}\end{bmatrix}$$

Figure 2: The f-structure of (2c) and (2d)

This problem is arose from the standard LFG assumption that the GF-POS mapping does not reflect linear order. Thus, the account of (1) based on f-structure cannot be applied to the contrast in (2c-d). The same holds of the examples in (3) and (4), which only differ with respect to the order of the conjuncts.

3 Generalization

Through these examples, we can observe one generalization: the head imposes its restriction (POS or number/person agreement) on an argument sufficiently close to it but not necessarily on an argument sufficiently far from it. In other words, only arguments which are sufficiently close to the head have to satisfy the syntactic requirements imposed by the head. This generalization covers (1)-(4) uniformly. For example, the head in (1a) imposes its requirements on *that*-clause since the place *that*-clause appears is in the range of constraints imposed by the head.

[1] We call *that*-clause as CP for convenience.

[2] For various alternative technical formulations of this proposal, see the references cited above.

On the other hand, *that*-clause in (1b), is sufficiently far from the head on which the grammatical restriction from the head is not imposed [3]. Likewise, in (2c), though the second conjunct is the complement of the head *of*, the head does not impose its constraints on it, since the second conjunct is far from the head and the syntactic requirement is not necessarily imposed.

This intuitive observation itself is not new, and it is already proposed in the literature (Sadock 1998; Moosally 1998) that the head agrees only with the nearer conjunct. However, this observational generalization has failed to be stated explicitly in their formalized theories; Sadock only mentions the observation, and Moosally only stipulates the agreement patterns. Moreover, their accounts cannot cover sentences such as (1). Also, in the unbounded dependencies accounts, non-LFG frameworks such as HPSG requires the syntactic category of the filler to match the requirement imposed by the selecting head. This problem may be resolved by revising their assumptions for the unbound dependencies. However, it is difficult for structure-based analysis to manipulate the syntactic mechanism since it would fail to capture the linear order effect which we have observed above, hence it would be difficult to account for (1) and (2)-(4) uniformly. This leads us to consider that this phenomenon is not of syntactic nature, which is supported by the following example support our intuition.

(5) a.?Ken was thinking of, (pause) that he was stupid

b. Ken was thinking of, by the way, that he was stupid.

The observation is that the insertion of a pause improves the acceptability of (5a), while the additional insertion of *by the way* makes sentence (5b) fully acceptable. On the standard assumption, a pause and *by the way* only affect real-time linear order, not syntax. But the observation is, at least intuitively, exactly the same as the one we found for (1)-(4); the syntactic head *of* fails to exert its constraints on its complement when the complement becomes further from it. Thus, syntactic accounts fail to capture our intuitive generalization.

Let us restate our informal generalization as in (6).

(6) The Linear Order Effect:

The syntactic requirement the head imposes on an argument is effective only to the extent that the argument is "sufficiently close" to the head in linear order.

In the next chapter, we attempt to model this observation.

4 The Memory-based Sentence Processing Model

What we can observe from example (5) is that the phenomena in question should not be accounted for by the syntactic resolution based on the word order or adjacency, but by the real-time processing.

We propose a processing model based on the assumption given in (6), which we call "memory-based sentence processing model".

(7) Memory-based Sentence Processing Model:

The syntactic information is deactivated on the following conditions:

(i) when the predicate-argument structure is assumed to have been constructed by the parser

(ii) when the processing complexity is too much in relation to the processing time

Condition (i) means that the syntactic information is fully deactivated when the semantic content is assumed to have been obtained. Likewise, condition (ii) means that the syntactic information is gradually deactivated by the passage of time. Note that the deactivation degree is not at the same from condition (i) to (ii). This difference arises from the assumption that the syntactic information is necessary for constructing the predicate-argument structure.

Also, this model is based on the general idea that syntactic information is needed only to obtain the semantic content, and that the capacity of working memory is severely limited. These lead us naturally to expect that the syntactic information is deactivated rapidly from working memory as soon as it has played the role of constructing the semantic content.[4]

When the processor takes in a word, it constructs a corresponding tree, and as the tree is successfully constructed, the processor predicts the forthcoming input on

[3] Note that semantic constraints are fully imposed.

[4] Takahashi and Ishikawa (2004) have assumed that the syntactic information is "deleted". However, considering backtracking, it seems that syntactic information which is stored in working memory should be deactivated gradually.

the look-ahead basis and expands the tree. When the inputted word is a head, the processor immediately constructs its complement node on a look-ahead basis and imposes various syntactic/semantic constraints on the complement, which has not been actually encountered yet. Note that this is the role of grammar. If the phrases with expected part-of-speech information appears and the predicate-argument content is completed, then the whole syntactic information of the sentence is deactivated rapidly and only the semantic structure is left.

Let us consider the case of complement coordination. As encountering the head, the parser constructs the complement node, on which the head imposes various constraints. When the parser is fed with the first conjunct, which satisfies the constraints, the syntactic information is deactivated rapidly since the parser assumes that thepredicate-argument structure is completed. However, as a conjunction such as "and" is inputted, the parser has to modify the complement tree, so as to dominate a newly constructed node for the second conjunct. When the second conjunct is actually inputted, the constraints imposed by the head have already lost their effects. Therefore, the sentence is acceptable even if the second conjunct fails to obey the syntactic constraints imposed by the head, provided that the semantic constraints are all satisfied.

5 Demonstrations

In this section we show how the memory-based model accounts for the sentences in question.

5.1 Topicalization

In our memory-based model, (1a) is unacceptable since a syntactic restriction is imposed on the argument immediately after it has been fully activated by the head. On the other hand, (1b) is acceptable since the syntactic information of the topicalized phrase is so deactivated that the incompatibility between the topicalized phrase and the head is made less influential even though the constraints from the head is imposed. Note that the semantic constraint is fully imposed on the topicalized phrase because the semantic content is not deactivated from working memory.

5.2 Complement Coordination

We assume that, when dealing with a constituent coordination structure with two conjuncts, the human parser initially constructs a structure containing only the first conjunct before reading the conjunction (*and*) and after that, it combines with the second conjunct; the structure is subsequently modified into a coordinate structure when the conjunction and the second conjunct are encountered (see Figure 3 for an illustration). This assumption is experimentally supported by Sturt and Lombardo (2005).

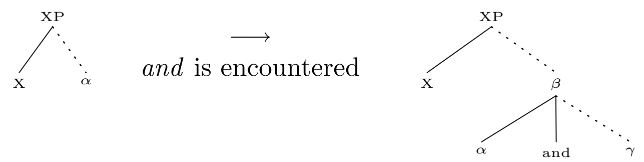

Figure 3: The initial partial tree and the reconstructed partial tree

With this assumption in hand, the pattern in (2) can be explained as follows.

Upon encountering *about*, the parser constructs its complement NP node. When the parser is fed with the first conjunct, which happens to be an NP (α in Figure 3), the syntactic information is fully deactivated since the parser (wrongly) assumes that predicate-argument structure is completed (the left figure in Figure 3).

5.3 Insertion

In our model, the contrasts in (5a-b) are due to the presence of a time interval between the head and the complement (the pause in (5a) and the inserted adjunct in (5b)). The difference between (5a) and (5b) is probably due to the processing complexity of an inserted overt phrase. An inserted phrase consumes processing resources more than a mere pause, a natural assumption about the nature of working memory. Since the syntactic constraints become more deactivated through processing an inserted phrase than an inserted pause, the acceptability of (5b) is better than (5a).

5.4 RNR

In our model the Right Node Raising examples (3a-b) in which the linear order seems to affect their acceptability status are accounted for as follows: in (3a), *about* and the "raised" phrase *that Mike was wrong* are adjacent, while in (3b) they are not. In our model, the constraints imposed by the head are loosened or deactivated as time passes by. Moreover, since the forces of constraints imposed by the first conjuncts in each sentence are more deactivated than those of the second conjunct, it is the second conjunct that contains the head, and is adjacent to the raised complement. The syntactic constraints of the first conjuncts in each sentence are lost by the time when the parser encounters the raised complement, and the constraints imposed by the second conjunct have to be fully obeyed. Therefore, (3a) is unacceptable while (3b) is acceptable.

6 The Grammar-Processor Interaction

In section 2, we have pointed out the problems that some linguistic phenomena fails to be accounted for only by syntax; rather, they should be accounted by the real-time processing. Still, we also consider that the nature of the parser affects the stage of processing specified by the grammar. This is because, in some cases, though the grammar attempts to construct the structure correctly, the parser may have produced only an insufficient because of the limits of the capacity of working memory and the efficiency of sentence processing. In this section, we argue the grammar-parser interaction, which is based on our model and combinatory categorial grammar (Steedman 2000). The main reason for adopting combinatory categorial grammar is its power of description: combinatory categorial grammar is not based on the structural assumptions which do not reflect the word order. Thus, combinatory categorial grammar can incrementally analyze sentences in linear order [5], in the actual order of human sentence processing.

As an example, we show how the complement coordination can be analyzed in terms of categorial grammar and our model incrementally.

When processing from left to right, in the same way as human sentence processing, combitonary categorial grammar stops its process when the first conjunct is encountered since the sentence is assumed to be completed, as shown in figure 4. ">B", ">" and "$< \Phi >$" is applications to calculate categories in combinatory categorial grammar:

(8) a. $X/Y \;\; Y \;\; \Rightarrow \;\; X \;\; (>)$

b. $X/Y \;\; Y/Z \;\; \Rightarrow \;\; X/Z \;\; (>\mathsf{B})$

c. $X \; CONJ \; X' \;\; \Rightarrow \;\; X'' \;\; (< \Phi >)$

However, this is an expected process, for in our model, we want the parser to finish the process once the predicate-argument structure is assumed to have been constructed. At this point, since the semantic content is obtained, the syntactic information is deactivated. However, the parser is betrayed; the conjunction *and* is encountered and the parser has to recognize that the process is not finished. At this point, the derivation is backtracked. However, since the previous syntactic information is deactivated, the category of *his girlfriend* is fully deactivated (*yp*) and the syntactic category of *John was thinking of* is set to S/xp. However, the semantic content ($\lambda x.is\ thinking\ of'x\ john'$)is already obtained. This category represents that the sentence is not ended, and the syntactic constraints including case checking are deactivated. Note that this is not due to the grammatical mechanism, but to the parser effect.

As the conjunction is encountered, the grammar requires that the syntactic category be the same one as the first conjunct. However, the category of the first conjunct is a variable, when the *that*-clause is encountered: *yp* unifies with S_{+cp}[6]. Likewise, *xp* unifies with S_{+cp}, resulting in acceptable when the semantic constraints are satisfied. We illustrate the process after the backtracking in figure 5.

7 Conclusion

We have pointed out the problem in analyzing a certain kind of sentence only in terms of grammar: we have argued that, it should be dealt with in terms of real-time processing, and proposed a memory-based sentence processing model, demonstrating how our model accounts for the problematic sentences uniformly. Then we have showen how combinatly categorial grammar analyzes the complement coordination.

However, there remain several problems. First, though we have assumed that the syntactic information is deacti-

[5] Combinatory categorial grammar itself does not assume that the sentences are parsed in the real word order. The process following word order is just one of its possible of derivations

[6] We use "S_{+cp}" for *that*-clause for convenience.

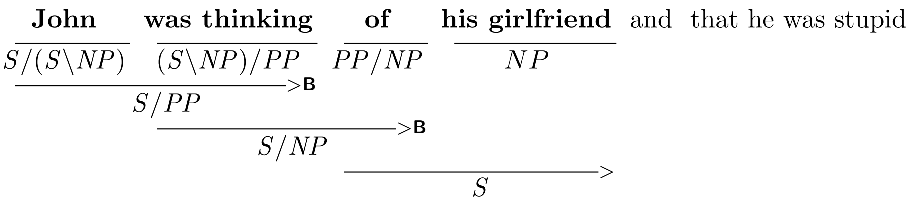

Figure 4: The partial process of (2c)

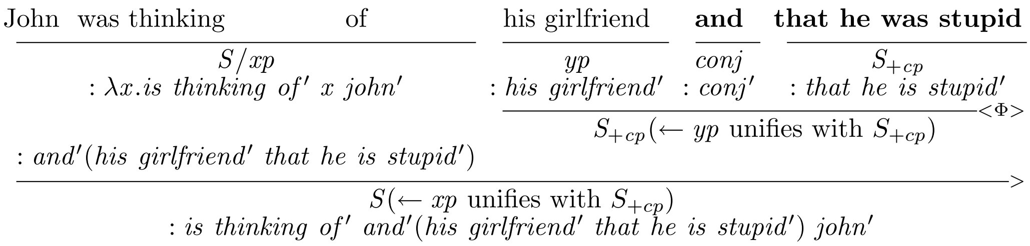

Figure 5: The syntactic process after the deactivation in (2c)

vated, it is not clear what kind of syntactic information is deactivated. Also, it has not been clarified that to what extent the information is deactivated. Another question is, when the pronouns and *that*-clause are coordinated, they are judged to be unacceptable even if the order of conjuncts are changed. Currently, our model cannot account for such examples. Moreover, we assumed that the parser predicts the forthcoming category on a look-ahead basis. However, we have not defined how the parser predicts the category. Narayanan and Jurafsky (2004) accounts for this problem in terms of Bayesian Theory. We agree with their assumption that statistical method can be used in the human sentence processing. To solve these problems, carrying out the psychological experiments is our future task.

Acknowledgement

We appreciate The Tohoku University 21st Century COE Program In Humanities (Stragetic Research and Education Center for an Integrated Approach to Language, Brain and Cognition; The LBC Research Center) and Center for Interdisciplinary Research (CIR, Tohoku University) for cooperation. Also, we thank to Kiyoshi Ishikawa (Hosei University) and Rui Otake (Tohoku University) for discussion.

References

Bresnan, Joan. 2000. *Lexical-Functional Syntax*. Oxford: Blackwell.

Falk, Yehuda N. 2001. *Lexical-Functional Grammar: An Introduction to Parallel Constraint-Based Syntax*. Stanford: CSLI Publications.

Hawkins, John. 1994. *Performance Theory of Order and Constotuency.* Cambridge: Cambridge University Press.

Kaplan, Ronald M., and Joan Bresnan. 1982. Lexical-Functional Grammar: A Formal System for Grammatical Representation. *The Mental Representation of Grammatical Relations*, ed. Joan Bresnan, 173–281. Cambridge, MA: MIT Press.

Kaplan, Ronald M., and Annie Zaenan 1989. Long-Distance Dependencies, Constituent Structure, and Functional Uncertainty. *Alternative Conception of Phrase Structure* , ed. Mark R. Bartin and Anthony S. Kroch, 17–42, Chicago: University of Chicago Press.

Kaplan, Ronald M. 1995. The Formal Architecture of Lexical-Functional Grammar. *Formal Issues in Lexical-Functional Grammar*, eds. Mary Dalrymple, Ronald M. Kaplan, John T. Maxwell and Annie Zaenen. 7–27. Standord: CSLI Publication.

Moosally, Michelle J. 1999. Subject and Object Coordination in Ndebele: An HPSG Analysis. *WCCFL 18* , eds. S.Bird, A Camnie, J. Haugen, P. Norquest, 379–392. MA: Cascadilla Press.

Narayanan Srini, and D. Jurafsky. 2004. Bayesian Model of Sentence Processing. unpublished manuscript, available on `http://www.icsi.berkeley.edu/~snarayan/`

Quirk, R. et al. 1995. *Longman Dictionary of Contemporary English, 3rd ed.* Harlow: Longman.

Sadock, Jerrold M. 1998. Grammatical Tension. *CLS 34: The Panels*, 179–198.

Sturt, Patrick, and Vincenzo Lombardo. 2005. Processing Coordinated Structures: Incrementality and Connectedness. *Cognitive Science* 29, 291–305.

Steedman, Mark *The Syntactic Processes.* MA: MIT Press

Takahashi, Kei, and Kiyoshi Ishikawa. 2004. An Adjacecy Constraint on Argument Selection. *Language, Information and Computation: Proceedings of the 18th Pacific Asia Conference*. eds. Hiroshi Masuichi, Tomoko Ohkuma, Kiyoshi Ishikawa, Yasunari Harada, and Kei Yoshimoto. 23–34. Tokyo: The Logico-Linguistic Society of Japan.