An Adjacency Constraint on Argument Selection *

Kei TAKAHASHI
The University of Tokyo
3-8-1 Komaba, Meguro-ku,
Tokyo 153-8902, JAPAN
kei-ta@phiz.c.u-tokyo.ac.jp

Kiyoshi ISHIKAWA
Hosei University
2-17-1 Fujimi, Chiyoda-ku,
Tokyo 102-8160, JAPAN
kiyoshi@i.hosei.ac.jp

Abstract

Instead of positing seperate syntactic mechanisms, we propose a single working meory mechanism that uniformly accounts for (i) a puzzle about topicalization pointed out and explained in the LFG literature, (ii) asymmetries in coordination structure observed and analyzed in various syntactic frameworks, and (iii) the effects of inserted phrases and pauses.

1 Introduction

As Chomsky (1957) made clear, natural language syntax cannot be described *solely* in terms of linear order (finite-state grammar); hierarchical structure plays a crucial role. However, that does not mean that linear order has no role to play in our accounts of native (un)acceptability judgments. To say the least, given that native judgments are results of real-time processing, it would rather be surprising that linear order had no effect on such judgments. Thys, some researchers (e.g. Hawkins 1994) have attempted to explain seemingly syntactic phenomena in terms of real-time processing.

In this paper we join such a recent trend by making a specific proposal about linear order effects and forumulate the proposal in terms of working memory. The structure of the paper is as follows. In Section 2, we illustrate the (un)acceptability judgments we intend to account for; we point out that the existing accounts fail to capture the observational generalization behind the constructions. In Section 3, we propose a specific linear order account, modeled in terms of working memory. In Section 4, we demonstrate that our proposal nicely accounts for the phenomena in question. In Section 5, we further examine the nature of our proposal by suggesting an account of a potential counterexample. In Section 6, we reflects on what the merits of our account is more precisely. Section 7 concludes the paper.

2 Data and Problems

Kaplan and Bresnan (1982) pointed out the contrast in (1), which poses a problem (at least) for a movement-based analysis of topicalization. An obvious explanation for the ungrammaticality of (1a) would be to assume that *of* cannot take a *that*-clause as its complement. This leads us t expect that topicalization of the *that*-clause does not alter the sentence's unacceptable status, an expectation surprisingly 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) crucially relies on the LFG assumption that complement selection is stated in terms of grammatical function (GF), instead of part of speech (POS); the linking of GF and POS is stated by a separate constraint. Putting technical details aside, the gist of the proposed account is that (i) while TOP (= topic) can be realized as a CP, OBJ (= object) cannot; OBJ can only be realized as an NP, (ii) *of* selects OBJ, and (iii) the syntactic relation in (1b) between *of* and the topicalized clause is stated in terms of GFs. (1a) is bad

* A very earlier version of the intuition underlying the theory described here was presented at the Japan Cognitive Science Society (JCSS) annual meeting in 2003. We thank the JCSS audience as well as Shûichi Yatabe (the University of Tokyo) for helpful comments, Mafuyu Kitahara (Waseda University) for an informal discussion, and Campbell Hore (Nara Institute of Science and Technology) for native judgments.

because, in violation of (i), a *that*-clause attempts to realize OBJ, the GB selected by *of*. In contrast, in (1b), the OBJ in question is not realized by an overt expression (no violation of (i)); indeed, the OBJ value is the f-structure of the *that*-clause, but since an f-structure contains no POS information, this does not violate (i).[1]

However, such accounts fail to predict the grammaticality differences in the following examples, which intuitively seem to be related to the pattern observed in (1). (4c–d is from Quirk et al. (1995, §10.41), cited from Yatabe(2004).)[2]

(2) a. John was thinking of [Mary].

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

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

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

(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.

For example, given that (2c–d) differ only in the order of the conjuncts (which are bracketed), they should have exactly the same f-structure, given the natural and standard LFG assumption that f-structure does neflect linear order. The GF-POS mapping is naturally not assumed to reflect linear order, either. Thus, an account of (1) based on f-structure and GB-POS mappings cannot be extended to the contrast in (2c–d). Similarly for (3)–(4), which all share the pattern that the good and bad examples only differ with respect to the order of the conjuncts.

Observationally, it seems to us, the generalization is this: the head imposes its restriction (POS or number/person agreement) on an argument near enough to it but not necessarily on an argument far enough from it. In other words, only those arguments near enough to the head in linear order have to satisfy the grammatical requirements imposed by the head. This intuitive generalization convers (1)–(4) uniformly.

Such an intuition itself is not necessarily new. For example, the agreement asymmetry in coordinate structure as seen in sentences such as (4) is a rather old observation, and it is already assumed in the literature (Sadock 1998; Moosally 1998; Yatabe 2004) that the head agrees only with nearer conjuncts.An exception is Johannessen (1998), who analyzes a coordinate structure as the maximal projection of the

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

[2](2c) is subsumed under this explanation only when (α) *Mary and that he was stupid* is analyzed as a constituent; if (β) *of Mary* is a constituent coordinated with the *that*-clause, (2c) would simply be an ordinary constituent coordination which poses no interesting problem. (We thank Shûichi Yatabe for pointing this out.)

However, consider (i) and (ii).

i Sally missed Japan so much she was thinking nostaligically of even election-vans and that rush-hour on the Yamanote-line was not so bad as people make out.

ii *? Sally missed Japan so much she was thinking nostalgically even that rush-hour on the Yamanote-line was not so bad as people make out and election-vans.

In (i), *even* modifies the NP and the *that*-clause, which indicates that the NP and the *that*-clause form a constituent, which functions as the complement of *of*, thereby demonstrating that (α) is a possible analysis. Further, the contrast between (i)–(ii) suggests that, even when the (β) analysis is forced, the linear order generalization is observed; the inversion of the conjuncts in (iii) degrades the grammaticality.

Of course, the badness of (ii) could be a result of the mere fact that the left conjunct is too long and hence places too much burden on working memory, independently of our specific working memory model to be developed below. However, even if we granted such an argument and assumed that (2c) is accepted only on the (β) parse, (2a–d) would then only fail to support our proposal; they would not refute it anyway.

conjunction, in which the specifier and the head agree (the Spec-Head agreement assumed in GB). In a language like English in which the specifier always precedes the head, then, the first conjunct should agree with the head, which agrees with its maximal projection in turn, which is directly selected by the head and hence should satisfies the head's requirements. In short, this predicts that, if there is an asymmetry in a coordinate structure, it is the first conjucts (but not the other conjuncts) that obey the syntactic requirements imposed by the head selecting the coordinate structure, *irrespective of word order between the head and the selected coordinate structure*. For example, it predicts the opposite pattern for (4). As far as we are aware, every other published work on coordinate structure asymmetries as well as our own informant judgments are in conflict with his prediction. However, this observational generalization has failed to be stated explicitly in their formalized theories; Sadock only mentions the observation, and Moosally and Yatabe only stipulate agreement patterns on a case-by-case basis. And their accounts per se are not meant to, and cannot, cover non-coordination sentences such as (1), on the other, thereby failing to capture the similarity between (1) and (2)–(4). It is not that they merely failed to notice the similarity and to point out that their generalization also covers (1). Any existing non-LFG account of unbounded dependencies (either in HPSG assumed by Moosally or Yatabe, or any other framework we are aware of) requires the syntactic category of the filler to match the requirement imposed on the complement by the selecting head (P in this case). Thus, unless one revises either one's treatment of unbounded dependency constructions (the syntactic approach) or one's assumption of the role played by syntax in native speaker judgments (the non-syntactic approach), their accounts cannot be extended to cover (1).

The syntactic approach would be an obvious course to take, but in this paper we dare take the non-syntactic approch. For one thing, manipulating syntactic mechanisms for unbounded dependency constructions in non-linear order terms, as in LFG, would fail to capture the observational generalization of linear order effects, and hence does not seem likely to succeed in giving a uniform treatment of (1) and (2)–(4). For another, the following examples lead us to suspect that the nature of the linear order effects is rather not syntactic:

(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 (5a), and the additional insertion of *by the way* makes the sentence fully acceptable (5b). On the standard assumption, a pause and *by the way* only affect real-time linear order, not syntax. But the observed pattern is, at least intuitively, exactly 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, a syntactic account fails to capture our intuitive generalization.

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

(6) The Linear Order Effects (observation):
The syntactic requirement the head imposes on an argument is effective only to the extent that the argument is "near enough" to the head in linear order.

Our task then is to clarify the notion of "near enough."

3 The Working Memory Model

3.1 The General Idea

The leading idea behind our overall approach is that the Linear Order Effects (6) is a result of the need for real-time processing (parsing or generation by a human agent).

Firstly, it is already observed that, after hearing a sentence, the overall form of the sentence is easily lost from memory, while the sementic content (in the sense of the predicate-argument structure) is retained (Sachs 1967).[3] Indeed, there has to be some time interval in order for the overall form to get lost from memory, but it at least suggests that the goal of parsing is to construct the semantic representation of the sentence; syntactic information is only a means to achieve that goal. Then, it would not be totally implausible to imagine that, while the overall form is retained in memory, more finer details of "syntactic"

[3]This information is due to Abe (1995).

pieces of information[4] such as the head's POS requirements on of the sentence is more easily lost from memory than the overall sentence form.

Secondly, real-time processing proceeds "from left to right." Given that syntactic constituency is rather orthogonal to the left-to-right linear arrangement, real-time sentence processing crucially relies on short-term memory, or working memory, an assumption already assumed widely in the psycholinguistic literature. Given that the capacity of working memory is severely limited, then, it would be rather natural, and at least computationally preferable, that useless information get "expelled" from working memory as soon as possible.

These two observations lead us to the hypothesis that the fine-grained syntactic information of a sub-sentential expression is "expelled" from working memory even before the end of the sentence. At the point when a given piece of syntactic information is "expelled" from working memory, that piece fails to exerts its influence on the processing of the remaining parts of the sentence, which amounts to saying that the relevant syntactic constraint loses its force at that point. Our idea is to account for the (un)acceptability judgments observed above with this hypothesis.

Although this is not the right place to discuss various models of working memory (or short term memory), we assume that some sort of activation model is correct. That is, to say that an item is in working memory is to say that that item is activated in a specific way. (That item might be an item previously stored in long term memory before processing the sentence in question or an item constructed on-line.) This assumption means that whether an item is in working memory or not is not a yes-no matter but rather a matter of degree; comletely "expelled" items (or those items not put in working memory in the first place) have the activation value of zero, completely active items have the full activation value, and there are also those items whose activation values are somewhere in between.

3.2 The Specific Details

Having illustrated the overall general idea, we now specify the details of our model, in a step by step fashion.

3.2.1 Adjacency

Assume that a pharse (maximal projection) P_1 (here, conceived as a node) is selected by a phrasal head P_2, which should come after P_1. Once the processor has succeeded in constructing P_1, P_2 can be immediately constructed, on a "look ahead" basis, before actually encountering those words that are (or turn out to be) parts of P_2. At this point, P_2 is constructed in working memory, with syntactic information fully consistent with the requirements imposed by P_1.

Next, assume that a lexical head L selects, and hence imposes some syntactic information on, a phrase P_2 that should come after L. In such a case, too, once the processor has processed L it can immediately construct P_2 on a "look ahead" basis, which we assume the processor does. Thus, at this point, P_2 is put in working memory, with syntactic information fully consistent with the requirements imposed by L.

Whichever is the case, the next incoming words should ideally be those which are parts of P_2; otherwise, the processor's expectation is betrayed, the attention is detracted from P_2, and hence, the syntactic information on P_2 is deactivated, given the limited capacity of working memory.

Thus, in both cases, P_2 should ideally be adjacent to P_1 or L; otherwise, the syntactic constraint C dictated by the grammar on P_2 will fail to exert its (full) effect on the processing of the sentence, where the (un)acceptability judgment is a result of processing. In other words, the "degree of adjacency" influences acceptability judgments (given that it makes sense of the "degree of adjacency"): the further P_2 is from P_1 or L, the less severe the violations of C.

3.2.2 Syntax vs. Semantics

Our general idea is that contents in working memory is activated only to the degree they are useful for the construction of the semantic content, here understood as the predicate-argument structure. Thus, what we mean by the term "syntactic" above should be understood as "whatever information is not part of the predicate-argument structure." For example, the phonological content is clearly not part of the predicate-argument structure and hence is "syntactic" in the sense used here. Similarly for the

[4] By the term "syntactic" here we mean whatever fine aspects of grammar that are not part of the final product of parsing: the semantic representation (predicate-argument structure). In the remainder of the paper, this term is used in this sense unless specified otherwise.

information concerning number and person agreement, even if one assumes that agreement information is part of semantics, not syntax.[5]

3.2.3 The Deactivation Degrees

We have thus far proposed that the "degree of adjacency" determines the degree of deactivation. Usually, the term "adjacency" is understood as a yes-no matter, with no gray zone. However, given that the observed adjacency effects are a result of the deactivation in working memory, it will be natural to talk about the "degree of adjacency."

Now, what does it mean for two items I_1 and I_2 not to be adjacent? It means that some other item I_3 intervenes between I_1 and I_2. Thus, our thinking leads to the conclusion that the degree of non-adjacency is just the degree to which the intervening item I_3 deactivates the syntactic information on I_2.

With all those backgrounds, the observed (un)acceptability judgments can be explained in a uniform manner if something like the following deactivation rate assignment is assumed, where 1.0 is the "fully activated" level, 0.0 is the fully deactivated (or "expelled") level, and the effects of the presence of intervening elements are additive:

- While processing I_3, where I_3 is a maximal projection argument, 1.0 is subtracted from the activation level of the syntactic information on I_1 and I_2.
- While processing I_3, where I_3 is a maximal projection adjunct, 0.3 is subtracted.
- While processing I_3, where I_3 is a pause, 0.3 is subtracted.
- If the result of a subtraction becomes less than zero, it counts as zero.

In the next section we demonstrate that this model successfully acounts for the observed (un)acceptability judgments in a uniform manner.

4 Demonstrations

In this section we demonstrate that the above model successfully accounts for the judgments in question in a uniform manner, but before proceeding, a word about the theory of grammar is in order.

A processor is a device that constructs linguistic representations using the knowledge of grammar. Thus, the precise behavior of a processor depends on the theory of grammar. In our demonstration, we have HPSG in mind as the theory of grammar. However, our reliance on the specific architecture of HPSG is minimal, and we believe that our model can be implemented on other grammatical frameworks alike.[6]

4.1 Coordination

The observed generalization is that the head agrees with the closest conjunct, where the term "agree" is meant to cover both number/person agreement and POS requirements. We examine the POS requirements on complements and the number/person agreement on subjects, in that order.

4.1.1 Compelement Coordination

The relevant examples (2a–d) are repeated here as (7):

(7) a. John was thinking of Mary.

b.*John was thinking of that he was stupid.

c. John was thinking of Mary and that he was stupid.

d.*John was thinking of that he was stupid and Mary.

We assume that *of* in these examples requires an NP complement, as indicated by (7b). The generalization is that, given a configuration of the form:

[5]There is good evidence that number agreement sometimes affects the propositional content of the sentence, a favorite topic for one of us (for example, see Ishikawa 1998). However, the question of whether number agreement is semantic or syntactic (in the usual sense of the term) is a thorny and complicated question. We avoid a discussion of it and simply assume that it is not part of the semantic content (the predicate-argument structure), an assumption widely held in the GB/MP literature.

[6]The crucial features of HPSG we reply upon will be noted in footnotes.

V $[_X\ Y$ and $Z\ ]$

only P_1, the first or closest conjunct, is required to obey the syntactic requirement imposed by V.[7]

Immediately encountering V, the processor constructs Y. At this point, the processor does not know yet[8] that the complement of V is a coordinated phrase, and hence, there is no point in positing X or Z. The incoming words are analyzed as parts of Y (Figure 1).

The syntactic requirements imposed by V on Y are fully active, and the POS of the first conjunct Y has to match the requirement by V. Hence the unacceptability of (7d).

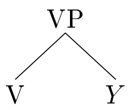

Figure 1: The initial partial tree

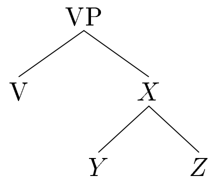

Figure 2: The modified partial tree

In contrast, upon encountering the conjunction (in this case, *and*), the processor has to modify the already constructed structure to something like Figure 2. X here is a newly constructed node here. However, at this point, 1.0 is subtracted from the activation level of the subcategorization information of V (=I_1), while processing the argument phrase Y (=I_3).[9] This means that the subcategorization requirement fails to be imposed on X and hence Z; hence the acceptability of (7b).

The examples in (3a–b) receive a similar account.

4.1.2 Subject Coordination

The relevant examples (4) are repeated here as (8):

(8) 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.

Here we ignore the presence of the expressions *either* and *or*.[10]

In (8a), I_2 is the VP node. First, assume that I_1 is the first conjunct, *she*. Immediately after processing *she*, the processor does not know yet that it gets coordinated with another NP. Thus, it constructs a VP node, with the third person singular agreement information. However, while processing the second conjunct *you* (=I_3), which is an argument, 1.0 is subtracted from the activation level of this agreement information on the VP node. Thus, third person singular agreement is not imposed on the VP node anymore. Also note that the agreement information on the first conjunct (=I_1) has also decayed at this point. On the other hand, at this point the mother node has to be constructed for the coordinate subject. The second conjunct has just been constructed, and hence, its second person agreement information is still active in working memory. Hence, the mother node for the coordinate subject bears the second person agreement specification, hence the second person agreement on the VP and hence the head V.

A similar account applies to (8b–c).[11]

[7] Here we follow HPSG and most other syntactic frameworks (but not LFG) in assuming that a lexical head imposes POS requirements on its complements.

[8] In this paper the processor is assumed to be a parser.

[9] Here we are using the term “argument” in the sense that it is a semantic argument. It may or may not be an argument in the syntactic sense, depending on one’s precise analysis of the syntax of coordination, which is beyond the scope of this paper.

[10] We discuss these expressions in a footnote, after giving the general idea of how our model works.

[11] Speaking precisely, this is not the case, since in (8b–c), we observe that agreement with the first conjuncts does not produce full unacceptability, which the above accounts predict.

We suspect that this complication is related to the presence of the expressions *either* and *or*. While the proper syntactic analysis of the *either ... or* construction is not at yet clear, one possible line of analysis would be one in terms of Construction Grammar, of which HPSG can be seen as an instance. If the general idea of Construction Grammar is assumed, it is plausible that the pharase *either X or Y* is stored in long term memory as an underspecified phrase, which is put in working memory as it is. This would significantly complicate the story, depending on whether such a construction counts as an argument or an adjunct. We suspect that the mild unacceptability in question is due to such complication, but we have to leave this issue to future research.

4.2 Topicalization

The relevant examples (1a–b) are repeated here as (9a–b):

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

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

In (9a), *of* is I_1, which constructs an NP node by "look ahead"; the subcategorization requirement of the P imposed on its complement is fully active. Thus, the processor expects an NP, an expectation betrayed here; hence the unacceptability.

In contrast, in (9b), things are radically different. Immediately after constructing the topic phrase node ($=I_1$), the processor constructs an S node by "look ahead." However, the processor next has to construct the subject NP, and then the VP node, and then the PP node, before reaching the P *of*. At least the NP and the PP are arguments. Thus, while constucting these nodes, the activation level of the topic phrase is fully deactivated and hence the part-of-speech information is no longer accessible. Hence the full acceptability of (9b).

4.3 Insertion

The relevant examples are repeated here:

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

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

In (5a), a pause ($=I_3$) in inserted between *of* ($=I_1$) and the node for the *that*-phrase ($=I_2$). According to our mode, when actually encountering the words comprising the *that*-clause, the syntactic information on I_1 and I_3 have changed from 1.0 to 0.7 (=1.0-0.3), hence the degraded status of (5a).

In (5b), there are three intervening elements, the inserted phrase *by the way* (which is an adjunct) and the two pauses surrounding it. Each element substract .3 from the activation level of I_1 and I_3, the additive result is .9 subtraction ($.9 = .3 times 3$), which we assume is large enough and can be equated with 1.0 subtraction.[12] This means that the syntactic requirement imposed by *of* is effectively erased from working memeory, and (5b) is correctly predicted to be fully acceptable.

5 Semantic Constraints

One interesting observation is provided by Bayer (1996), who observes that *despite* rejects a *that*-clause complement, even when the clause is a non-first conjunct seperated from it by some other conjunct(s).

(11) a. Despite [LaToya's intransigence], Michel signed the contract. (NP)

b. Despite [the fact [that all the musicians quit]], Michel signed the contract. (NP)

c.*Despite [that all the musician quit], Michel signed the contracs. (*that*-clause)

d.*Despite [LaToya's intransigence] and [that all the musician quit], Michel signed the contract. (NP conjoined with *that*-clause)

This suggests that the constraint that *despite* imposes on its complement does not follow the pattern dictated by our model.[13] We mention this possible counterexample in order to illustrate more precisely the nature of our model. We point out that this is not a counterexample to our model if a semantic analysis along the line suggested here is accepted.

[12]We could alter the degree of subtraction in the previous section from .3 to $\frac{1}{3}$ without affecting the story at all, in which case the prediction is 1.0 subtraction.

[13]This leads one to wonder how the requirements imposed by *despite* manifest in such extraction constructions as topicalization. However, as Huddlston, Pullum and Peterson (2002) point out, *despite* does not allow extraction of its complement in the first place:

i *Here is a list of the objectors, [that they went ahead [despite __]].

and thus we cannot examine whether it follows the patter or not in extraction constructions.

5.1 Contentfull and Contentless Prepositions

It is rather an old observation in the HPSG literature that English Prepositions have two kinds: those predicative Preposition with semantic content (P[+PRD] in the usual HPSG treatment), and those which only serve as a sort of case marker and hence are without semantic content (P[-PRD]). For example, in (12), the first *for* is a mere "case marker" with no semantic content, whereas the second *for* does have semantic content and says that the complement's referent is the person who is to benefit from Ken's behavior:

(12) Ken is waiting for Naomi for me.

The observation that *despite* does not follow the pattern dictated by our model and exhibited by *of* in (1) and (2) is, we claim, due to the fact that the former is a contentful P[+PRD] while the latter is a contentless P[-PRD].

Let us assume that only contentful heads can (but do not necessarily have to) impose semantic restrictions on their complements. Thus, *despite* imposes semantic restrictions on its complement, while *of* in question does not. But what kinds of semantic restrictions are imposed by *despite*?

5.2 Selection Restrictions and the Classifications of Events

A classical argument against the possibility of giving a semantic account of the so-called Complex NP Constraint is that, while a mere *that*-clause allows extration from an element inside it, a supposedly synonymous complex NP *the fact that ...* does not. Putting aside the issue of how observations that motivate the Complex NP Constraint should be accounted for, this argument is clearly invalid. Various nouns other than *fact* can take a *that*-clause complement, as in:

(13) the fact/rumour/claim/... that ...

Thus, the natural assumption is that a bare *that*-clause is neutral as to whether it expresses a fact, a rumour, or a claim, etc. (cf. Asher 1993). Let us call the distinction between a fact, a rumour, a claim, etc, **P-SORT** (short for "proposition sort"). Hence, a bare *that*-clause is not synonymous, for they differ with respect to P-SORT.

The above observation can be expressed in at least the following two ways:

- The P-SORT of a bare *that*-clause is underspecified.
- The P-SORT of a bare *that*-clause is *nil*, a special sort that is distinct from those sorts such as *fact*, *rumour*, *claim*, etc.

If things are formulated in HPSG, in which case P-SORT will be adopted as a semantic feature whose possible values include *fact* (and possibly *nil*), the former analysis means that the value can unify with any sort (it only says that the grammar imposes no constraint on the value):

(14) $\left[\text{SYNSEM}\,|\,\text{CONT}\,|\,\text{P-SORT} \quad [\;]\right]$

while the latter means that the grammar constrains the value to be distinct from any other sort such as *fact*, *rumour*, or *claim*:

(15) $\left[\text{SYNSEM}\,|\,\text{CONT}\,|\,\text{P-SORT} \quad \text{nil}\right]$

Both analyses could be formulated in HPSG, but if we choose the latter (i.e. the relevant portion of the feature structure of a *that*-clause is as described in (15) and assume that *despite* selectionally restricts its complement to be [SYNSEM|CONT|PSORT $\neg$ nil], our model makes the correct predictions, where we assume that the lexical entry for *fact* is partially described in Figure 3:

When it takes a *that*-clause complement, the clause unifies with $\boxed{2}$, which structure-shares the P-SOA value $\boxed{3}$ with the head noun *fact*. In other words, the propositional content is "inherited" from the *that*-clause complement to the head noun *fact*, and hence to the whole NP *the fact that ...*. On the other hand, the *that*-clause complement and the head noun *fact* have distinct P-SORT values: *fact* and *nil*. That is, the propositional sort is not inherited. Thus, while a bare *that*-clause bears the [P-SORT nil] specification and hence rejected by *despite*, such an NP as *the fact that ...* bears the [P-SORT fact]

$$
\left[\text{SYNSEM}\,|\,\text{LOC}\ \left[\begin{array}{ll}
\text{CAT} & \left[\begin{array}{ll}\text{HEAD} & \text{N}\\ \text{COMPS} & \langle \boxed{2} \rangle\end{array}\right]\\
\text{CONT} & \left[\begin{array}{ll}\text{P-SORT} & \mathit{fact}\\ \text{P-SOA} & \boxed{3}\end{array}\right]\\
\text{ARG-ST} & \langle \boxed{1}, \boxed{2}\left[\text{SYNSEM}\,|\,\text{LOC}\,|\,\text{CONT}\left[\begin{array}{ll}\text{P-SORT} & \mathit{nil}\\ \text{P-SOA} & \boxed{3}\end{array}\right]\right] \rangle
\end{array}\right]\right]
$$

Figure 3: A partial description of the lexical entry of *fact*

specification, which successfully unifies with the [P-SORT ¬nil] requirement imposed by *despite*, and hence can successfully combines with it.

The assumption that *despite* requires a non-nil complement is a reflection of the intuition that the complement of *despite* has to denote something which is already established and somehow presupposed. For example, in (16a), the fact that the content of the *that*-clause is a rumour/claim etc. is already established and presupposed:

(16) a. despite the rumour/claim that ...

b. *despite that ...

In contrast, in (16b), only the P-SOA (the propositional content) is stated. The propositional content is expressed even by an imperative, and hence, by itself does not establish anything. This is the cause of the unacceptability.

In short, we suggest that the real generalization is not that *despite* rejects a non-NP complement but rather that it rejects a complement which resists a factive (or presuppositional) reading. Indeed, many NPs allow a factive interpretation in this context, and hence, the inherent factivitity is not in question. Rather, it rejects an expression that cannot be interpreted as an already established fact. Our choice of assigning the *nil* as the P-SORT value to a bare *that*-clause, instead of simply underspecifying it, reflects this observation.

5.3 The Survival of Semantic Constraints

Remember the intuitive idea behind our model: the requirement imposed by a linguistic item on another item decays in working memory to the extent that it has finished playing its role for the constrution of the semantic content. This means that semantic constraints should survive, irrespective of word order; they survive even after the processing of the whole sentence is well finished.

This in turn means that the semantic constraint imposed by *despite* survives, no matter what the linear order is. That is, it requires a complement whose P-SORT value is not *nil*. This is why a bare *that*-clause is not allowed even as a second conjunct, given that a semantic requirement imposed on a coordinate structure distributes to each of its conjunct. [14]

---

[14] Shûichi Yatabe (p.c.) points out that a bare *that*-clause is sometimes interpreted as a fact:

(i) a. [That John was drunk at that time] caused traffic accident.

b. [The fact [that John was drunk at that time]] caused traffic accident.

c. [That John won the lotto] surprised his girlfriend.

d. [The rumor [that John won the lotto]] surprised his girlfriend.

In (ia) and (ic), it does make sense that the subject is interpreted as a fact. We can point out that the verbs require the subjects to be interpreted as facts. On the other hand, the subjects in (ib) and (id) presuppose that the content of the *that*-clause is a fact or rumor. That such a factivitity presupposition is operative is suggested by the degraded status of (ii):

ii ?[That John might won the lotto] surprised his girlfriend.

However, this suggests that a bare *that*-clause sometimes does allow a "fact" interpretation, i.e. as a [P-SORT fact] phrase, an observation rather in conflict with the account illustrated in the main text. A discussion of this conflict is left for future research.

6 Discussion

Our model successfully gives a uniform account of the observed (un)acceptability judgments, but the success partially owes to the price we have paid: the stipulation of numerical degrees of deactivation. We have decided the numerical degrees so that our model would give the correct predictions for (5a–b). To the extent that the numerical specifications have no independent evidence, our account of (5a–b) could be called a mere paraphrase of the observed degrees of unacceptability, or a translation of the observation to a description in terms of working memory.

However, while a mere observation makes no further empirical predications, our "paraphrase" or "translation" definitely does. Indeed, we have not confirmed the predictions at this point. However, a model or theory which makes unexpected empirical predictions indicates a future research agenda and hence is fruitful, we assume.

In this section, we first point out that our account makes at least one further, independent predication. And then we conclude this section with a somewhat speculative remark.

6.1 Phonological Priming

The conceptual justification for our assumption that "syntactic" information decays in working memory is that the goal of processing a sentence is to produce a semantic representation, and hence, it is costly to keep an item active in working memory even after it has become useless and is something that a human processor is likely to avoid. This predicts that phonological information becomes deactivated too (possibly even before purely syntactic information becomes deactivated).

This prediction could be empirically tested. A word is known to cause various priming effects. For example, assume you assign a lexical decision task to the subjects in a controlled experimental setting[15] and measured the response time (RT). The RT for the test word (called the target) becomes faster if you had presented another word (called the prime) that has a close semantic relation with the test word beforehand, than if you hadn't, in which case you have observed a(n) (indirect) semantic priming effect. A similar priming effect, called a(n) (indirect) phonological priming effect, is observed when the prime and the target resemble phonologically to each other. The usual assumption is that such a priming effect is observed only while the relevant semantic or phonological information is active in working memory. Thus, if our model is correct, we naturally expect not only that phonological priming effects disappear more rapidly than semantic priming effects but also that the disappearance rates of phonological priming effects correlate with the degree of the failure for the relevant syntactic constraints to manifest their effects.

Although we have not actually confirmed or disconfirmed these predictions, the numerical assignments in our model are not mere paraphrases of the observations to the exent that they make such independent predictions.

6.2 Timing vs. Overt Lingusitic Materials

In visual and auditory perceptions, the stimulus is often occluded by irrelevant noise, in which case the ability to restore those parts missing from the stimulus (or to reconstruct the whole shape of the original stimulus based on the observed parts) is crucial for the survival of the agent. However, two cases should be distinguished: those cases in which some parts are simply missing, and those cases which the missing parts are replaced with some "noise." Automatic perceptual restoration occurs only in the latter case. For example, a sequence of discrete pure tones seperated by silent intervals is perceived as just that, a series of pure tones and silent intervals (no restoration). However, if an appropriate white noise is inserted into each silent interval, suddenly it is perceived as one continuous pure tone with white noises imposed on it (restoration).[16]

The distinction between the two cases can be understood as a distinction between the absence of the expected material (a pure tone) and the presence of something unexpected (a white noise). Thus understood, the distinction begins to sound relevant to our discussion of the linear order effects. In sentence processing, the distinction corresponds to the distinction between (A) the absence of the expected expression and (B) the presence of an unexpected expression. But what counts as an "(un)expected" expression?

[15] A lexical decision task is a task to decide whether the presented stimulus "word" is a real word in your native language or a fake word that in fact is not part of your language.

[16] See (and hear) Bregman (1990) and Bregman and Ahad (1996) for demonstrations and discussions of various visual and auditory restorations, including this one. The well-known Phoneme Restoration Effect can be seen as a subspecies of this.

In our present model, the presence of a pause and the presence of an adjunct are assumed to have the same numerical effect, as opposed to the presence of an argument. If a pause does not count as an "unexpected expression," no parallelism could be pointed out betwen our present model and the above observation in visual and auditory perception. If we want to seek for some non-trivial parallelism, we would have to assume that a pause does count as an "unexpted expression."

Of course, we are not saying that the observations in perpetual psychology and our own discussion of the linear order effects are completely parallel. (Definitely we are not dealing with "restoration.") However, if our present model is on the right track, it suggests that a pause has a similar status as an overt linguistic expression such as an argument or an adjunct.

Note that we are not advocating that a pause should be conceived as a linguistic expression in the narrow sense (an empty category); a pause is only a silent time interval. Rather, the suggestion is that a time interval does make a difference, which is known to be the case in phonetic peception.

For example, suppose that *can* in (17a) is pronounced with a stress on it, in which case the vowel becomes indistinguishable from that in *can't* in (17b) ([æ] in American English, as opposed to a schwa).

(17) a. I cán do it. [kæn]

b. I can't do it. [kænt]

Crucially, [t] in (17b) is not released and hence fails to exhibit a burst, and because of the presence of [d] immediately following it, its presence or absence cannot be distinguished on the basis of the formant transition of the following vowel. Thus, in such a case, (17a–b) cannot be distinguished on the basis of the presence or absence of [t]. Rather, they are perceived differently because of the presence of a silent interval between [n] and [d] in (17b) that is not found in (17a).[17]

If our model is on the right track, it suggests that the presence of a mere time interval exhibits an effect not only in phonetic perception but also in (un)acceptability judgments. If our reasoning is correct, the reason is obvious: both are something done in real-time. Real-time processing is, by definition, crucially time-dependent, and it would be rather surprising that information concerning time would exhibit no effect in human speech behavior in real time, subspecies of which are phonetic perception and (un)acceptability judgments. This suggests that lingusits should pay more attention to real-time processing than they have done.[18]

7 Conclusion

In this paper we have proposed a specific working memory model to account for the (un)acceptablity judgments. The merit of our proposal argued for in this paper is that it gives a *uniform* account of the observed judgements, instead of formulating a seperate (syntactic) mechanism to account for each observation. Indeed, its further prediction is yet to be (dis)confirmed, and various technical details are yet to be refined. However, if it is on the right track, our approach suggests that linguists should pay more attention to the fact that (un)acceptability judgments are a result of real-time processing and hence is not immune to the working of the human processor. A seemingly grammatical phenomena is not necessarily grammatical.

References

Abe, Jun-ichi. 1995. Bun-no Rikai. *Ninti Sinrigaku 3: Gengo*, ed. Yukio Otu, Tokyo: University of Tokyo Press, 159–171.

Asher, Nicholas. 1993. *Reference to Abstract Objects in Discourse*. Dordrecht: Kluwer.

Bayer, Samuel. 1996. The Coordination of Unlike Categories. *Language* 72(3): 579–616.

Bregman, Albert S. 1990 *Auditory Scene Analysis: The Perceptual Organization of Sound*. Cambridge, MA: MIT Press.

[17] We regret that we cannot point to the original literature; we owe the story to an informal introductory lecture on phonetics given to the second author by Takehiko Makino (p.c.) at a dinner table.

[18] We thank Mafuyu Kitahara for pointing out that a pause is usually assumed to be an indication that the speaker is trying to figure out what (s)he should say. One may or may not call this "lingusitic," but if it is not "lingusitic," our model could be interpreted as a claim that such non-lingustic information does have a syntactic effect in real-time processing, a claim about the interaction between the linguistic and the non-linguistic.

Bregman, Albert S., and Pierre A. Ahad. 1996. *Demonstrations of Auditory Scene Analysis : The Perceptual Organization of Sound*. Cambridge, MA: MIT Press. (Audio CD)

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

Chomsky, Noam. 1957. *Syntactic Structure*. The Hague: Mouton.

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

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

Huddlson, Rodney, Geoffrey K. Pullum and Peter Peterson. 2002. *The Cambridge Grammar of English Language*. Cambridge: Cambridge University Press.

Ishikawa, Kiyoshi. 1998. On Plural Anaphora. *Language, Information and Computation: Proceedings of the 12th Pacific Asia Conference*. ed. Jin Guo, Kim Teng Lua, and Jie Xu. 184–194. Singapore: Chinese and Oriental Languages Information Processing Society.

Johannessen, Jonne B. 1998. *Coordination*. Oxford: Oxford 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* , eds. 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.

Pollard, Carl, and Ivan A. Sag. 1994. *Head-driven Phrase Structure Grammar*. Chicago: Chicago University Press.

Reape, Mike. 1994. Domain Union and Word Order Variation in German. *German in Head-driven Phrase Structure Grammar* , eds. John Nerborne, Klaus Netter, and Carl Pollard, 151–198. Stanford: CSLI Publications.

Sachs, J. S. 1967. Recognition Memory for Syntactic and Semantic Aspects of Connected Discourse. *Perception and Psychophysics* 2: 437–442.

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

Shûichi, Yatabe. 2004. A Comprehensive Theory of Coordination of Unlikes, to appear in *Proceedings of the 11th International Conference on Head-Driven Phrase Structure Grammar* ed. Stefan Muller, Stanford: CSLI Publications