Syntactic Underspecification in a Neuro-Cognitive Model of Human Sentence Processing

Kei Yoshimoto (kei@linguist.jp)
Shigeru Sato (satos@mail.tains.tohoku.ac.jp)
Graduate School of International Cultural Studies, Tohoku University
Kawauchi 41, Aoba-ku, Sendai, 980-8576 Japan

Abstract
A formal linguistic model is proposed which simulates human neuro-cognitive behaviors in processing German sentences with and without unambiguous case markings. The grammar formalism is crucially based on syntactic underspecification as a key concept. Keywords: electroencephalography, human sentence processing, Dynamic Syntax, underspecification

Introduction
Despite experimental data being accumulated by non-invasive brain activity measurement methods such as electroencephalography (EEG) and functional magnetic resonance imaging (fMRI), no attempts have ever been made to model human sentence comprehension based on those data from a unified formal perspective. Friederici’s (2002) three phase model still remains seminal and non-formal. In this paper, we propose a formal sentence processing system which simulates different event-related potential (ERP) elicitation between sentences with and without unambiguous case marking. As a formal linguistic theory we adopt Dynamic Syntax (DS; Kempson et al. 2001), which enables incremental sentence processing by underspecifying syntactic trees.

ERPs in Sentence Reanalysis
Bornkessel et al. (2003) and Schlesewsky and Bornkessel (2004) observed different ERPs, which have been associated with language processing in the brain, in the reanalysis of German head-final subordinate sentences with full morphological case marking and those with ambiguous case marking.

In comparing active verbs with an agentive subject like the one in (1a) and object-experiencer verbs which denotes the dative object person experiencing something as in (1b), both sentences fully case-specified, Bornkessel and the colleagues observed a remarkable ERP, an early parietal positivity P345, for object-experiencer verbs in contrast to active verbs. However, for sentences such as (2a) and (2b) in which proper name and plural NPs are ambiguous between nominative, dative, and accusative, they found no difference. Furthermore, in an experiment on subject-initial and object-initial structures in case-ambiguous sentences such as (3a) and (3b), an N400, i.e., an ERP usually considered to reflect semantic violation, was observed for object-initial sentences in contrast to subject-initial ones.

The results show that different incremental tree buildings are involved with case-specified and case-ambiguous trees: while unambiguously case-marked sentences are processed by fixing a syntactic tree step by step, ambiguous case NPs are left unspecified in terms of their syntactic positions until the verb is read in. It is also supposed that a sentence with insufficient information is interpreted temporarily by default, both in syntax and semantics. The NP der Mönch unambiguously marked as nominative in (1b) at first undergoes a thematic role interpretation as agent, which is canceled later by the object-experiencer verb demanding a less oblique thematic role (i.e., ‘experiencer’) for the dative than for the nominative NP. For sentences with ambiguous case markings (e.g., (2a,b)) no fixed tree is built before the verb, thus free from the need to reanalyze. Furthermore, while the first NP Maria in (3a,b) is at first not integrated into the whole tree, a temporary decision about their syntactic position as subject is made, which may later prove to be incompatible with the verb inflection.

Underspecification in DS
Dynamic Syntax is a formal linguistic theory which enables by establishing modal relationships between syntactic nodes incremental processing of head-final clauses such as German subordinate clauses and Japanese main sentences. Based on this mechanism, we propose to introduce reanalysis and default parsing which are motivated by the EEG observation cited above. The modified version of the theory, which is a formal extension of Friederici’s (2002) model at the same time, is outlined as follows:

Iterate the execution of Steps 1 and 2 for each word from left to right until the end of the sentence:

(1) a. daß der Mönch dem Bischof folgt
b. daß der Mönch dem Bischof gefällt

(2) a. daß Maria Lehrerinnen folgt
b. daß Maria Lehrerinnen gefällt

(3) a. daß Maria Lehrerinnen folgt
b. daß Maria Lehrerinnen folgen

223
Step 1: A word is identified from the lexicon based on the word category.
Step 2: (i) A new syntactic-semantic representation is formed based on the lexical information and grammatical rules by extending the representation built so far.
(ii) For a word with insufficient morphological syntactic information, its relationship to the root node is left underspecified. A default syntactic position is assigned to the word, which is sanctioned as a final one if no inconsistencies are found.
(iii) To a word with a fixed syntactic position, a semantic role is assigned by default.

In executing the whole processing including Steps 1 and 2, if pieces of information do not match, processing is restarted back at the position where the cause of the mismatch arose. Both such backtracking and overwriting of default specifications are assumed to overload the processing and engender each characteristic ERP.

When the nominative NP *der Mönch* (monk) is read in parsing (1b), the least oblique thematic role ‘proto-agent’ is assigned to it by Step 2 (iii). The verb *gefüllt* (pleases), by contrary, requires that the dative NP be an experiencer (i.e., semantically less oblique) and the nominative be a patient (i.e., semantically more oblique). As in this case, when an object-experiencer verb is inserted clause-finally, the default interpretation <subject=proto-agent, ind-obj=beneficiary> which has been formed so far is overwritten by the case frame <subject=patient, ind-obj=experiencer> subcategorized for by the verb (see Fig. 1). We assume that the revision of the default specification demands excessive processing, causing the ERP.

Fig. 1: Default case role overwriting in (1b)

In contrast, the processing of a sentence with ambiguous case marking like (2a,b) leaves its subject, object, and verb syntactically underspecified by Step 2 (ii), and finally the relationship between them is settled based on the lexical/morphological information provided by the NPs and verb. Here, there is no need of overwriting semantic roles.

For ambiguously case marked sentences, a default parsing is made in parallel with underspecification, as stated in Step 2 (ii) above. In processing (3b), before reading the verb, its NPs and V remain underspecified as to its relationship to the root node (*the root node is somewhere above’) as indicated by the dotted branches in Fig. 2. In parallel, default parsing is made incrementally which gives assumptions on the syntactic positions of the nodes, as the solid branches in Fig. 2 show. In this case, the subject and object positions are assigned by default to *Maria* and *Lehrerinnen* (teachers), respectively. When the verb *folgen* (follow-PLU) is read in, however, the default analysis turns out to be inappropriate owing to the agreement mismatch (see Fig. 2). Then the default parsing is restarted at *Maria*, which is correctly given an object position this time. This backtracking in the default syntactic processing is assumed to engender another kind of ERP, i.e., an N400.

Fig. 2: Agreement mismatch in default parsing of (3b)

Conclusion
We have shown that the EEG data in the human processing of German sentences with and without unambiguous case markings can be accounted for appropriately based on syntactic underspecification. It is theoretically of interest whether the model can be applied to Japanese, which as a typical head-final language necessitates underspecified sentence processing. An experiment on this issue is now in progress using EEG and fMRI simultaneously.

Acknowledgments
The research was supported by the Tohoku University 21st Century COE Program in Humanities and by the Tohoku University CIR Grant for Research Program in Computational Neurolinguistics.

References