

## Clarifying the concept workspace, revising Merge to MERGE, and identifying consequences

Hisatsugu Kitahara (Institute of Cultural and Linguistic Studies, Keio University)

Yoshihito Dobashi (Faculty of Humanities, Niigata University)

Toru Ishii (School of Arts and Letters, Meiji University)

Nobu Goto (Faculty of Business Administration, Toyo University)

Takanori Nakashima (Graduate School, Faculty of Arts and Letters, Tohoku University)

The strong minimalist thesis SMT takes the computational system  $C_{HL}$  for human language to be a "perfect system," meeting the interface conditions in a way that satisfies third factor principles (see among others, Chomsky (1995, 2005)). Under SMT, the combinatorial operation of the generative procedure, which generates hierarchical structures, is expected to be very simple. Presumably, the simplest possible formulation is a set-formation device that takes  $X$  and  $Y$ , and forms  $\{X, Y\}$ , which has come to be called Merge (see Chomsky (2004, 2005)). And unless stipulated, we expect the following two instances of Merge: External Merge, when  $X$  and  $Y$  are distinct, and Internal Merge, when  $X$  is part of  $Y$ .

Does this formulation of Merge make sense? In his most recent work (in a series of lectures in Arizona, and particularly Reading from 2017), Chomsky notes that there is a problem. Merge has been sitting within a framework, left somewhat vague. That is, to form  $\{XP, YP\}$  (e.g. NP-VP), Merge must be able to construct syntactic objects SOs in parallel (meaning independently), and bring them together somewhere. This presupposes that there is a workspace, in which operations such as Merge are being carried out. But then what is the workspace  $WS$ ? This question hasn't been properly answered, and fixing it does have consequences. Chomsky revises Merge to MERGE, an operation on  $WS$ , not particular SO, and takes  $WS$  to be the stage of the derivation at any given point. This revision raises all sorts of questions, and Chomsky et al. (2017) explore some of them. For example, how does a derivation  $D$  terminate? They propose that  $D$  "may (but need not) terminate whenever  $WS$  contains a single object; if it terminates in any other situation, no coherent interpretation can be assigned." Or how does MERGE work? Chomsky et al. (2017) ask, "does  $MERGE(X, Y)$  add  $\{X, Y\}$  to  $WS = [X, Y]$  (where  $X, Y$  are LIs or complex elements), yielding  $WS' = [X, Y, \{X, Y\}]$ ? Or does it rather replace  $X$  and  $Y$  in  $WS$  with  $\{X, Y\}$ , yielding  $WS' = [\{X, Y\}]$  (as assumed in Chomsky 1995:243)?" They point out that "a workspace  $WS' = [X, Y, \{X, Y\}]$  derived by  $MERGE(X, Y)$  would not ensure that subsequent rules can apply in a determinate fashion: any rule referencing  $X$  or  $Y$  would ambiguously refer to the individual objects  $X, Y$  or to the terms of  $K = \{X, Y\}$ ."

In this workshop, following Chomsky et al. (2017), we would like to explore fundamental questions such as these, and identify some immediate consequences, drawing data from English and Japanese. We begin with a brief introduction, which reviews background assumptions and new proposals in Chomsky's recent work, and then present three papers: The first paper argues that a terminated derivation  $TD$  corresponds to an interpretive unit, given that the notion of  $WS$  makes it possible to explicitly determine in what situation a derivation may terminate, with reference to the interfaces (cf. termination by Numeration/Lexical Array (Chomsky 1995, 2000)). It argues that a  $TD$  is matched with a relatively large interpretive unit in the semantic and phonological components. Specifically, it shows that Intonational Phrase, which is often said to tie in with a semantic unit like speech act, can be recast in

terms of TD, without recourse to the direct phonology-semantics interactions. The second paper explores consequences of MERGE and Determinacy. Chomsky et al. (2017) propose seven desiderata that the computational operations of I-language including MERGE are subject to. One of the desiderata is Determinacy, which bans an ambiguity of rule application. If structural description for a rule holds for WS where there are more than one accessible copies, its structural change is not unique. This paper argues that such an ambiguous application of a rule violates Determinacy. It is shown that Determinacy provides us with a unified account of various phenomena like the subject condition, the *that-t* effects, the ban against vacuous topicalization, the freezing effects with topics, and improper movement. Finally, the third paper discusses how adjuncts can be analyzed within the framework outlined by Chomsky et al. (2017). It proposes that removal of SOs from WS may (but need not) take place when MERGE maps WS to WS', but it follows that removal of adjuncts takes place when Transfer applies upon the completion of each phase. Given this proposal, Determinacy prohibits any rule-application to those SOs remaining in WS; hence, the adjuncts, being multiple copies, constitute opaque (syntactically inapplicable) domain for rules, thereby explaining the adjunct condition effects, the argument/adjunct reconstruction asymmetries, and the backward anaphora constraint.

The workshop aims at gathering and discussing various issues related to these three papers, along with other issues discussed in Chomsky et al. (2017), hoping to gain new insights on the nature of C<sub>HL</sub>, in particular, how MERGE operates on WS. It also includes the introduction (serving as a mini tutorial) on recent development in minimalist syntax, which makes this workshop accessible to students and researchers from all levels.

#### Selected References

- Chomsky, Noam. 1995. *The Minimalist Program*. Cambridge, MA: MIT Press.
- Chomsky, Noam. 2000. Minimalist inquiries: The framework. In *Step by Step*, ed. by R. Martin, D. Michaels and J. Uriagereka, 89-155. Cambridge MA: MIT Press.
- Chomsky, Noam. 2004. Beyond explanatory adequacy. In *Structures and beyond: The cartography of syntactic structures*, Vol. 3, ed. by Adriana Belletti, 104-131, Oxford University Press, Oxford.
- Chomsky, Noam. 2005. Three factors in language design, *Linguistic Inquiry* 36, 1-22.
- Chomsky, Noam, Ángel J. Gallego, and Dennis Ott. 2017. Generative grammar and the faculty of language: Insights, questions, and challenges. Unpublished manuscript. lingbuzz/003507.

# Workshop: Using treebanks for linguistic research

## Organizer and speakers:

Yusuke Kubota (University of Tsukuba), organizer

Misato Ido (NINJAL)

Ayaka Suzuki (NINJAL)

Nobuyoshi Miyoshi (University of Tsukuba)

## Workshop description:

Although corpus-based study has become fairly common in descriptive grammar research (Miyake 2017), in theoretical research based on linguistic theories such as generative grammar, the use of corpora is still at an early exploratory stage. In particular, many important issues such as the methodology for using real attested data in (theoretical) linguistic research and practical issues that working researchers should be aware of when dealing with attested data have not been discussed adequately. (See Meurers (2005) for a seminal work that discusses various issues in this domain.) One reason for this somewhat unfortunate situation is that corpora representing detailed grammatical information are not easily available (Ogawa et al. 2016). In theoretical syntax and semantics, the main objects of research are the structures of sentences and the meanings of words and sentences for which structural information is essential. For this reason, it is difficult to utilize a corpus for full-scale research unless one can search for examples based on such detailed grammatical information in the corpus.

In view of the above situation, a new type of corpus called the NINJAL Parsed Corpus of Modern Japanese (NPCMJ) is currently being developed at the National Institute of Japanese Language and Linguistics. The NPCMJ corpus is a treebank (i.e. a syntactically annotated corpus) primarily designed as a resource for linguistic research. It contains fine-grained syntactic information indispensable for serious linguistic research in the above-mentioned sense. It is expected that this new

corpus will facilitate a new type of linguistic research and will ultimately bring theoretical linguistics research into a new stage as a true empirical science. This workshop introduces three case studies of linguistic research using the NPCMJ corpus, which, though preliminary, already corroborate this expectation. In theoretical linguistics research, the method of collecting data based solely on introspection of a small number of native speakers has been predominant. However, as discussed in detail in the three papers presented in this workshop, this traditional approach can be fruitfully complemented in various ways with a more data-oriented type of approach based on large scale attested data. The goal of the present workshop is to consider the possibilities and challenges of linguistics as a true empirical science through introducing concrete linguistic case studies using the NPCMJ corpus.

### **Abstracts of papers:**

**Paper 1:** Misato Ido (NINJAL) ‘Licensing of adverbial NPIs beyond clause boundaries’

As a case study using the NPCMJ corpus, this presentation examines a phenomenon in which adverbial negative polarity items (NPIs) in Japanese are licensed by negation beyond clause boundaries. In particular, I aim to show in this case study that using corpus data contributes not only to descriptive linguistics but also to theoretical linguistics. It is known that Japanese NPIs must normally appear in the same clause as negation unlike English NPIs such as *any*. However, Kudo (2000) and Matsui (2003), among others, pointed out that adverbial NPIs can be licensed beyond clause boundaries in copular sentences with predicative nominals. This presentation demonstrates that the rich syntactic annotation of NPCMJ makes it possible to extract attested data from the corpus verifying the phenomenon pointed out by Matsui. Furthermore, a qualitative analysis of the search results allows us to identify another group of adverbs that are similar to NPIs, which, together with NPI adverbs, point to a hitherto unnoticed generalization. Specifically, we identify three types of adverbs from the search results from the corpus: *mettani*-type (‘rarely’) NPI adverbs; *sonnani*-type (‘so much’) NPI adverbs; and *syottyuu*-type (‘quite often’) non-NPI adverbs. I will

show that these three types of adverbs have different types of meaning and distribution respectively. These syntactic and semantic features bring up a new research question pertaining to the syntax-semantics interface: despite showing different meanings and distributions, why do these three groups of adverbs occur commonly in noun modifier clauses for predicative nominals?

**Paper 2:** Ayaka Suzuki (NINJAL) ‘Temporal interpretations and tense forms of the *toki*-clause’

In this presentation, I argue that observations obtained from the NPCMJ corpus contribute to the investigation of the interpretive mechanism of the Japanese *toki*-clause. Specifically, an analysis using a parsed-corpus enables us to reexamine the assumptions that have been presupposed in previous studies, and to discover an analytical viewpoint that has been overlooked. Previous studies on the *toki*-clause have focused on the combinations of tense forms in the subordinate clause (SC) and the matrix clause (MC) (Mihara and Hamada 1996; Funahashi 2006; Oshima 2011), and hypothesized that if SC and MC have different tense forms, the tense in SC is interpreted relative to MC, and if they have the same tense form, the tense in SC is interpreted relative to the speech time. However, the patterns reflected in the attested data obtained from the NPCMJ corpus lead to a reexamination of this hypothesis. The key factor that affects the interpretation is whether the SC tense is the *ru*-form (i.e. the non-past form) or the *ta*-form (i.e. the past form) rather than whether the MC and SC tense forms are the same or different. Moreover, we can find counterexamples for previous studies’ hypothesis when SC has the *ru*-form. In order to give an explanation to these observations, we have to take the difference between the *ru*-form and the *ta*-form into consideration and assume that the fundamental meaning of the morpheme *toki* is to encode the notion of ‘simultaneity’. I show that ‘simultaneity’ of *toki* restricts the possible interpretation. The proposed analysis derives the descriptive generalization about the possible combinations of tense forms noted in the previous literature as a side effect of this fundamental meaning of the morpheme *toki*, similar in spirit to the account of Kaufmann and Miyachi (2011).

**Paper 3:** Nobuyoshi Miyoshi (University of Tsukuba) ‘Distribution of non-restrictive adnominal clause in different text genres’

Adnominal clauses in Japanese are known to have both restrictive and non-restrictive uses. The restrictive use functions as denoting a proper subset of the set of objects identified by the head noun, and the non-restrictive use adds some information to the head noun. Some previous studies have already referred to discourse functions and semantic classifications of the non-restrictive use. For example, according to Masuoka (1995), the two main functions of the non-restrictive use are ‘adding information to the head noun’ and ‘adding information to the matrix clause’. The latter is further subdivided into ‘contrast’, ‘sequence’, ‘reason/cause’, and ‘accompanying circumstance’. However, the function of the non-restrictive use in actual discourses has not been studied in adequate detail in the previous literature. The present study attempts to clarify the distribution and function of the non-restrictive use by closely examining its distributions in different text genres using the NPCMJ corpus. In the present survey, the following preliminary results were obtained. First, in expository or reporting texts such as newspaper, more instances of ‘adding information to the head noun’ were found than instances of ‘adding information to the matrix clause’. A possible explanation for this tendency is that discourse-new proper nouns and technical terms frequently occur in this type of text. Second, in novels, though the non-restrictive use was not as frequent, instances of the non-restrictive use in various semantic functions (including different subclasses of ‘adding information to the matrix clause’) were found. This is arguably due to the fact that novels differ from other text genres such as newspapers in exploiting a wider range of rhetorical strategies in discourse, such as scene setting and establishing causal relationships among events.

## References

Funahashi, Mizuki. 2006. Tokisetsu-no kaishaku-ni-kansuru goyooronteki koosatsu (Pragmatic factors in the interpretation of *toki*-clauses). *Nihongo Bunpoo* 6(1):106–121.

- Kaufmann, Stefan and Misa Miyachi. 2011. On the temporal interpretation of Japanese temporal clauses. *Journal of East Asian Linguistics* 20(1):33–76.
- Kudo, Mayumi. 2000. Hitei-no hyoogen (Expressions of negation). In S. Kinsui, M. Kudo, and Y. Numata, eds., *Toki, Hitei to Toritate (Tense, Negation and Focusing)*. Tokyo: Iwanami Shoten.
- Masuoka, Takashi. 1995. Rentaisetsu no hyogen to shumeishi no shudaisei (Expression of adnominal clause and topicality of head noun). In T. Masuoka, H. Noda, and Y. Numata, eds., *Nihongo no Shudai to Toritate (Topic and Toritate in Japanese)*, 139–157. Tokyo: Kurosio Publishers.
- Matsui, Haruko. 2003. Hitei-kyokusei-koomoku-no setsu-o koe-ta koo (Negative polarity agreement across clause boundaries). *Nihongo Bunpoo* 3(1):117–126.
- Meurers, W. Detmar. 2005. On the use of electronic corpora for theoretical linguistics: Case studies from the syntax of German. *Lingua* 115(11):1619–1639.
- Mihara, Ken-ichi and Miwa Hamada. 1996. Rentai-syuusyoku-gata-hukushi-setsu-no setsu-to tensu (Tenses in noun-modifying adverbial clauses). *Nihongo/Nihon-Bunka-Kenkyuu* 6:31–41.
- Miyake, Tomohiro. 2017. Bunpoosei-handan-ni-motozuku kenkyuu-no kanoosei (On the potential of grammaticality judgment as a reliable research method). *Nihongo Bunpoo* 17(2):3–19.
- Ogawa, Yoshiki, Akiko Nagano, and Akira Kikuchi. 2016. Gaikan: Gengo-henka/hen'i-no kenkyuu-to koopasu (Overview: Language change/variation and corpora). In Y. Ogawa, A. Nagano, and A. Kikuchi, eds., *Koopasu-kara Wakaru Gengo-henka/hen'i-to Gengoriron (Discovering Language Change and Language Variation from Corpora)*, 1–28. Tokyo: Kaitakusha.
- Oshima, David Y. 2011. On the interpretation of *toki*-clause: Beyond the absolute/relative dichotomy. *Journal of East Asian Linguistics* 20(1):1–32.