Linguistic Studies of Dialogu

Takao GUNJI, Michinao MATSUI, and Akira OHTANI
Department of Language and Information Sciences
Graduate School of Language and Culture, Osaka University
1-8 Machikaneyama-cho, Toyonaka, OSAKA 560, JAPAN
e-mail: gunji@lisa.lang.osaka-u.ac.jp
matui@lisa.lang.osaka-u.ac.jp
ohtani@lisa.lang.osaka-u.ac.jp

In this study, we have explored the relationship between the syntactic and semantic components of the Japanese grammar, on one hand, and the morphological and phonological components, on the other.

We have adopted the *constraint-based grammar* approach, which assumes that various kinds of information are shared and cross-referenced by many components of the grammar. The advantage of constraint-based grammar approach, as opposed to *derivation-based grammar*, such as transformational grammar, is that the grammar can express mutually constraining relationships among linguistic modules, which simultaneously constrain one another, and that no directionality in processing needs to be assumed or stipulated. Its grammatical descriptions generally take the form of statements about constraints among feature values, and about sharing of partial information represented by these features. In short, these are constraints on representations, distinct from those on derivations such as transformations (i.e., destructive restructuring such as transformations).

Since speech sounds are the interface of the language faculty in the human mind to the outside world, the relationship between the underlying representation and the surface representation in speech production and that in speech perception are the complete mirror image of each other. This property of speech helps us to obtain nondirectional representation of language, which is one of the advantages of constraint-based grammar from a computational point of view. In this sense, we have developed a constraint-based theory of Japanese phonology.

Two of the important concepts in constraint-based grammar are (1) *feature structures* and (2) *unification* among them. In our approach, two morpho-phonological features are used: the **morph** feature and the **phon** feature. The **morph** feature, which roughly corresponds to the underlying representation of speech sounds, specifies the phonological elements (roughly, phonemes, cf. Charette 1991) and the linear ordering among them. The **phon** feature, on the other hand, roughly corresponds to the surface representation and forms a mora-based phonological structure. These features can be represented by finite state automata (cf. Karttunen 1993, Bird and Klein 1994).

We have also investigated the relationship between the two major components in the grammar, which we call *tectogrammatical* and *phenogrammatical* levels, borrowing the terminology by Dowty (to appear). The former determines the semantic and syntactic information, while the latter morphological and phonological information. Coupled with this approach, our constraint-based feature system can flexibly represent the correspondence between the tectogrammatical information, on the one hand, and the phenogrammatical information, on the other in the lexicon.

In the first year, we have conducted some preliminary study concerning morphological and phonological relationships of verbs (verb stems) and adjacent auxiliary elements. As a result, we have demonstrated that many of apparent syntactic phenomena, such as word order variation (scrambling) and honorification, can be adequately treated mainly within the morphological and phonological component, i.e., the phenogrammatical level, provided information in other components, particularly in the tectogrammatical level, are freely accessible.

Another result obtained in the first year is that we were able to give a fairly complete description of the phonological shape of the verb stems and following auxiliary elements when the verb appears in highly inflected forms such as in the euphonic form ('onbinkei').

Our system in part draws on the idea of optimality. In this approach, grammatical constraints are designed to overgenerate possible alternatives for a representation of a linguistic information. An independent component, which is sensitive to universal and language-specific cost-inducing constraints, will help to choose the most likely candidate.

We have also developed a formal system of treating accent patterns, which were be further extended in the following years.

In the second and third years, we have tried to resolve several apparent problems in a purely constraint-based approach to phonology. One of the problems concerns itself with the unification of unconstrained feature structures: when two phonological features do not have a one-to-one correspondence to each other, which often happens in phonological phenomena, such a structure would result in violation of the constraint on unification. Hence, some constraints must exist to effectively prohibit such a violation. In our approach, we assume four constraints for this: the *Autosegmental Association Constraint* (AAC), the *Default Value Constraint* (DVC), *Strengthening Constraint* (SC), and the *Weakening Constraint* (WC). These constraints play important roles in word formation.

The AAC is originally proposed in the autosegmental theory (Goldsmith 1990). This constraint requires roughly that a phonological feature must correspond to another as much as possible. Thus, if there are more than one possible features that can correspond to a feature X, then the information carried by X is associated with all the features that correspond to X. In effect, the AAC produces long vowels, double consonants, euphonic forms (*onbinkei*), tone patterns of words, among others.

The DVC assigns a 'default' value to the features that is given no information. The 'default' value for the vowel is defined by the sonority hierarchy, and hence, in the case of original Japanese words, it is /i/ and in the case of Sine-Japanese words and loan words it is /u/. It plays an important role in the formation of loan words. The 'default' value for the consonant, on the other hand, is assumed to be /r/. That is, the feature **C** (for consonant) in the **morph** feature that has no corresponding phonological element is realized as /r/ in the **phon** feature. This hypothesis explains many interesting phenomena concerning /r/ in Japanese.

The SC 'puts' an additional mora when there is redundancy in the **morph** feature. That is, when the number of phonological structures is smaller than the number of morphological elements, an additional mora is realized in the value of **phon**. For instance, since Japanese verb /jom/ has 'one' mora structure, the final consonant /b/ is redundant. Thus, if this verb needs to be present as the root form, then the SC can give one more structure to make the values of the **phon** feature unify all of **morph** features.

On the other hand, the WC 'deletes' redundant values in the **morph** or **phon** feature.

For instance, when a consonant-chain with no **morph** elements exists (e.g., /ki·C·Ca·ri/), the WC 'deletes' the second mora structure (and the DVC gives the default consonant value /r/ to the C's) to complete unification. In this case, the final phonetic form of /ki·C·Ca·ri/ is [ki·ra·ri].

An important aspect in our approach is that we allow a variety of candidates for solutions, determining the 'optimal' one based on the concept of 'cost' associated with the constraints mentioned above. This idea is similar to that in the optimality theory. In this approach, grammatical constraints are deliberately designed to 'overgenerate' possible alternatives for a representation of the given linguistic information. An independent component, which is sensitive to universal and language-specific cost-inducing constraints, will help to determine the most likely candidate(s). According to this idea, we expect there to be more than one candidates for a phonological feature that corresponds to a morphological feature; the candidate that has the minimal degree of violation becomes the ultimate solution.

For instance, in producing a sentence, the **phon** feature can freely has any feature value. This means that many candidates can co-exist. But a candidate that keeps better correspondence between the **morph** and the **phon** features is of low cost and can be the ultimate solution. In another case, e.g., word formation, the unification of phonological structures plays an important role in the evaluation of cost.

The study reported here mainly discusses phonological variations (e.g., allophones, double consonants, etc.), word formation systems, and morpho-phonological interactions (especially verb-inflections in Japanese). We have also developed a formal system for representing accent patterns, which will be further extended in the future.

We have also developed a theory of complex predicates for which the phenogrammatical representation is essentially a flatter 'monoclausal' structure and the tectogrammatical representation is basically a 'biclausal' structure involving embedding of a verb phrase. This approach can explain both semantic and syntactic integrity, on one hand, and morphological and phonological integrity, on the other.

Thus, in our approach, the structure of a language signs are divided into two major levels: tectogrammatical level and phenogrammatical level, of which our study for the past three years investigated the relationship between the two levels, on one hand, and the internal detailed structure of the latter level, on the other.

References

Bird, S. and Ellison, M. (1994). One-level phonology: autosegmental representations and rules as finite automata. *Computational Linguistics*, 20.

Charette, M. (1991). *Conditions on Phonological Government*, Cambridge University Press, Cambridge.

Dowty, D. R. (to appear). Towards a minimalist theory of syntactic structure. In Sijtsma, W. and van Horck, A. (eds.), (to appear). *Discontinuous Constituency, Proceedings of the Tilburg Conference on Discontinuous Constituency*. Mouton de Gruyter, Berlin.

Goldsmith, J. A. (1990). *Autosegmental & Metrical Phonology*, Basil Blackwell, London.

Karttunen, L. (1993). Finite-state constraints. In J. Goldsmith (ed.), *The Last Phonological Rule*, University of Chicago Press, Chicago.

Prince, A. and Smolensky, P. (1993). *Optimality Theory: Constraint Interaction in Generative Grammar*, Technical Report 2, Center for Cognitive Science, Rutgers University.

Keywords : constraint-based grammar, morphology, phonology, autosegmental theory, finite-state phonology, constraint-based phonology, optimality theory, phonological feature structures adjacency, scrambling, causative, honorification