

Deriving Lexical Stratification from an Exemplar Model: The case of Japanese compounds

Past analysis of the phonological alternations involved in compounding Japanese stems have relied on reference to the lexical strata of the stems involved (McCrawly 1968, Kurison 2000, Ito and Mester 1995). The key observation is that Sino-Japanese stems, when compounded, undergo different processes than do Yamato Japanese stems. Reference to lexical strata in phonology, however, is certainly not limited to compounds in Japanese. The reason that *create* can be *creation*, but *break* can not be *breaktion* is said to be that Latinate stems must affix to Latinate roots (Spencer 1991). Certainly learners of a language can not be thought to have detailed etymological information at their disposal in learning the productive rules of their grammar, and yet, the domain of application of certain morpho-phonological alternations can be sufficiently described by invoking reference to the etymology of the constituents involved. In this talk, I will report on the use of an analogical model (Skousen 1989, 1995, 2002) to show that what appears to be lexical stratification based on word origin can be derived from non-local neighborhood effects in an exemplar model sensitive to sub-segmental (feature) specifications.

Recently, a number of researchers have turned their attention to phonological processes that interact with word frequency (Bybee 2000, Pierrehumbert 2001, 2002). In order to model these phenomena, they have applied a schema-abstraction model of memory to the storage and retrieval of words in the lexicon. The exemplar-based models store each experience with a word as an individual token and are capable of extracting patterns that develop over the course of language use. A critical characteristic of these models is their ability to categorize similarities across tokens based on phonological, or even phonetic, criteria. It is this aspect of the model that I invoke to explain the distribution of two distinct phonological processes involved in compounding Japanese stems.

The model I use calculates the lexical neighborhood of novel inputs based upon their phonological similarity to words already stored in the lexicon. It then categorizes the novel compound based upon similarities to other compounds in the neighborhood with assignment weighted to favor the influence of the closest neighbors. Other researchers have demonstrated the psychological reality of such neighborhoods calling on these effects to predict priming in processing tasks (Vitevitch and Luce 1998, Luce and Pisoni 1998) and probabilistic phonotactics (Vitevitch et al 1997, Vitevitch and Luce 1999). These researchers have calculated lexical neighborhoods by taking a neighbor to be any lexeme that differs by just one segment. My approach differs slightly from this tradition in that I expand the size of the neighborhood by allowing any member of the lexicon that shares a feature with the input to participate in categorization. This opens up the lexicon to generalizations that may be obscured on the segmental level and accounts for the distribution of phonological rules in Japanese that correspond to Sino-Japanese compounds and Yamato Japanese compounds, respectively.

Analysis based on both rule-based (Chomsky and Halle 1968, McCrawly 1968) and Optimality Theoretic (Kuriso 2000, Coetzee and Pater 2005) frameworks have crucially

relied on lexical stratification to account for sound patterns in natural language. The model presented here offers a theoretically neutral mechanism to derive this stipulation from aspects of lexical representation independently necessary to explain phonological alterations.

References

- Chomsky, Noam and Morris Halle (1968) *The Sound Pattern of English*, Harper and Row, New York.
- Coetzee, Andries and Pater, Joe. 2005. Gradient phonotactics in Muna and Optimality Theory. Ms, University of Massachusetts, Amherst, and University of Michigan.
- Derwing, Bruce and Royal Skousen (1994). Productivity and the English Past Tense: Testing Skousen's analogy model. In S.D. Lima, R.L. Corrigan, & G.K. Iverson (Eds.), *The Reality of Linguistic Rules* (pp.193-218). Amsterdam: John Benjamin.
- Ito, Junko and Armin Mester (1995) "Japanese Phonology." in J. Goldsmith (ed.). *The Handbook of Phonological Theory*, Blackwell Publishers, Cambridge, Massachusetts, pp. 817-838.
- Ito, Junko and Armin Mester (1996) "Stem and Word in Sino-Japanese," in T. Otake and A. Cutler (eds.), *Phonological Structure and Language Processing: Cross-Linguistic Studies*, Mouton de Gruyter, Berlin and New York, pp. 13-44
- Kuriso, Kazutaka. (2000) Richness of the Base and Roof Fusion in Sino-Japanese. *Journal of East Asian Linguistics* 9, 147-185.
- Luce, P.A. and Pisoni D.B. (1998). Recognizing spoken words: The neighborhood activation model. *Ear and Hearing*, 19, 1-36.
- McCrawly, James (1968) *The Phonological Component of a Grammar of Japanese*, Mouton, The Hague.
- Skousen, Royal (1989). *Analogical modeling of Language*. Dordrecht: Kluwer Academic Publishers.
- Skousen, Royal (1995). Analogy: a non-rule alternative to neural networks. *Rivista Di Linguistica*, 7, 213-231.
- Skousen, Royal (2002). "An overview of Analogical Modeling." in R. Skousen, D. Lonsdale, D. Parkinson (eds.), *Analogical Modeling: An exemplar-based approach to language*, pp.11-26.
- Spencer. A. 1991 *Morphological theory*. Oxford: Blackwell
- Vitevitch, M. S., Luce, P. A., Charles-Luce, J., and Kemmerer, D. (1997). Phonotactics and syllable stress: Implications for the processing of spoken nonsense words. *Language and Speech*, 40, 47-62.
- Vitevitch, M. S. and Luce, P. A. (1998). When words compete: Levels of processing in spoken word recognition. *Psychological Science*, 9, 325-329.