

Bayesian Learning of Lexical Classes in Japanese

Takashi Morita

October 26, 2015

Abstract

Some analyses of Japanese morphophonology assume four lexical classes (Table 1) (Ito and Mester, 1995; Fukazawa, 1998; Ito and Mester, 1999, 2009). The classification is based on two phonotactic constraints (OCP(voi) and *NC) and one morphophonological alternation (compound-voicing). This classification faces a learnability problem: the phonotactic constraints do not provide any deterministic cue for the existence of multiple classes. For instance, SJ2 and Foreign are distinguished by OCP(voi), but SJ2 is a subset of Foreign in terms of their potential range. Accordingly, the union of SJ2 and Foreign is equal to Foreign, and thus, the existence of SJ2 will be masked.

Given the difficulty of deterministic learning, this paper investigates Bayesian learning of the lexical classes in Japanese. In particular, I will examine whether a Bayesian learner can identify the correct four classes among the logically possible eight classes based on the three morphophonological properties (Table 2). One interesting discovery is that the Bayesian learning completely fails if it tries to assign (the most probable) classes to words simultaneously while identifying the classes. In other words, the Bayesian learner has to first identify which kinds of classes exist in data without class assignments, though this strategy is only successful under certain conditions, either.

	Native	Sino-Japanese 1 (SJ1)	Sino-Japanese 2 (SJ2)	Foreign
OCP(voi) requirement	Yes	Yes	Yes	No
Compound-voicing (<i>rendaku</i>)	Yes	Yes	No	No
*NC requirement	Yes	No	No	No

Table 1: Lexical classes in Japanese (Ito and Mester, 2009).

References

- Fukazawa, Haruka. 1998. Multiple input-output faithfulness relations in Japanese. Rutgers Optimality Archive ROA-260-0698.
- Ito, Junko, and Armin Mester. 1995. The core-periphery structure of the lexicon and constraints on reranking. In *Papers in Optimality Theory iii: University*

Class ID		OCP(voi)	*NÇ	Compound voicing
1	Native	1	1	1
2	—	1	1	0
3	SJ1	1	0	1
4	SJ2	1	0	0
5	—	0	1	1
6	—	0	1	0
7	—	0	0	1
8	Foreign	0	0	0

Table 2: Logically possible classes with respect to three morphophonological properties (OCP(voi), *NÇ, Compound voicing). Active constraints for a class are marked as “1” and inactive constraints as “0”.

of massachusetts occasional papers 32, ed. Jill Beckman, Suzanne Urbanczyk, and Laura Walsh, 181–210. Amherst, MA: GLSA Publications.

Ito, Junko, and Armin Mester. 1999. The phonological lexicon. In *The handbook of japanese linguistics*, ed. Natsuko Tsujimura, 62–100. Oxford: Blackwell.

Ito, Junko, and Armin Mester. 2009. Lexical classes in phonology. In *Handbook of Japanese linguistics*, ed. Shigeru Miyagawa and Mamoru Saito, chapter 4, 84–106. Oxford: Oxford University Press.