Studying Phonological Variation Using the Corpus of Spontaneous Japanese

Shin-Ichiro Sano
Keio University

In this talk, I present several case studies of phonological variation in Japanese using the *Corpus of Spontaneous Japanese* (NINJAL 2012, hereafter CSJ). Specifically, the studies examine the effect OCP(voice), which bans multiple occurrences of [+voice] in proximity of each other (aka. Lyman's Law, Itô and Mester 1986) on the patterns of two phonological processes in Japanese: 1) devoicing voiced obstruent geminates (geminate devoicing), and 2) rendaku. The results show that the patterns of these two processes are controlled by the same mechanism.

Geminate devoicing is the process where voiced obstruent geminates (hereafter voiced geminates) in Japanese loanwords, such as [bb], [dd], and [gg] devoice, resulting in [ff], [tt], and [kk] (Kuroda 1965, among others). When loanwords with word-final voiced obstruents are borrowed into Japanese, the original voiced singletons are realized as voiced geminates followed by an epenthetic vowel (e.g. 'dog' => /doggu/). However, voiced geminates are traditionally prohibited in Japanese phonology. Therefore, these segments undergo some phonological processes to fit into the well-formed segmental configurations in Japanese. The most common phonological process involved is devoicing, as in /doggu/ 'dog' => [dokku].

Rendaku (aka. sequential voicing), one of the well-known and well-studied morphophonological processes in Japanese, refers to the voicing of initial voiceless obstruents of the second member of morphologically derived words (most likely compounds) (Vance 1979, 1980). For example, when the two nouns /hosi/ 'star' and /sora/ 'sky' are morphologically concatenated, the initial consonant /s/ of the second noun becomes voiced, producing /hosizora/ 'starry sky.'

However, not all loanwords/compounds, undergo geminate devoicing/rendaku; that is, these processes show variability. Prior work has identified various kinds of lexical, phonological and morphological factors that affect the applicability of the processes (Itô and Mester et seq.; Nishimura 2003, 2006; Kawahara 2006 et seq.; Pater 2009; Vance 1979 et seq.;). This study focuses on the effect of OCP(voice), and the interactions between OCP(voice) and other factors, such as the distance between trigger/blocker and the potential undergoer of the processes, and the number of triggers/blockers.

An exhaustive search of the data in the CSJ resulted in 1) a total of 1,617 tokens of potential undergoer of geminate devoicing, of which 472 (29%) showed devoicing, and 2) a total of 1,153 tokens of potential undergoers of rendaku, of which 709 (59.4%) showed rendaku. Whether a segment underwent the processes or not was coded based on the annotation in the CSJ.

The results of the examination of geminate devoicing and rendaku show that 1) the ratio of the application of the processes is higher/lower when there is a trigger/blocker (voiced obstruent), showing that OCP(voice) is active, 2) geminate devoicing and rendaku is more likely/less likely to apply when the distance between the trigger/blocker and the target is closer, and 3) geminate devoicing and rendaku is more likely/less likely to apply when there are more triggers/blockers. These observed patterns are consistent with cross-linguistic tendencies (Locality: Itô & Mester 2003; Frisch 2004; Frisch et al. 2004; Ihara et al. 2009, Number: Tesar 2007; Kawahara 2011). I conclude that OCP(voice), distance, and number are a common source of variation that affects the distribution of both geminate devoicing and rendaku in Japanese.