A probabilistic model of Japanese accent

Yu Tanaka
Doshisha University

Ito and Mester (2016; henceforth IM16) have given a formal account of accentuation in Japanese stems. Although their foot-based analysis well accounts for the general patterns, yet to be known is whether it still holds where actual accent data which contain numerous exceptions are concerned. To fill the gap, this study proposes a probabilistic model based on IM16's grammar and tests it against large-scale loanword data.

A Maximum Entropy Harmonic Grammar (MaxEnt HG; see Goldwater and Johnson 2003) was trained on the accent patterns of 1,053 loanwords (mostly taken from Mutsukawa 2009) with the constraints proposed by IM16. Additionally, FAITH-OO(Prominence) was included in the model, in order to capture the fact that not a few English loans retain their original prominence (stress) location (e.g. *ánimaru* 'ánimal'). The overall constraint weights learned by the model (Table 1; key constraints boldfaced) roughly resemble the rankings in the "Weak Antepenultimate System," the subgrammar in IM16 with the ranking NONFINALITY(FT´), RIGHTMOST » WORDACCENT » INITIALFOOT. This indicates that a probabilistic grammar, when facing real data with considerable variation and influence of source languages, reaches a similar conclusion as IM16.

Table 1

Constraint	Weight
MINWDACC	12.09
NOLAPSE	3.71
NONFIN(FT´)	**2.64**
FOOTBIN	2.13
FAITH(Prom)	2.06
NONFIN(σ)	2.00
RIGHTMOST	**1.70**
PARSE-σ	1.33
WORDACC	**0.92**
INITFOOT	**0.00**

The model now also makes probabilistic predictions about the accent patterns of novel words. A closer look at the results raises some interesting issues. Although the predictions seem to generally match Japanese speakers' intuitions on how nonce words should be pronounced, the rate of unaccentedness is somewhat higher than what is expected by IM16. This is possibly due to the fact that the training data actually contain quite a lot of unaccented words which may be rooted in socio-linguistic factors (see Inoue 1998). Also, the MaxEnt model does not always project the exact same foot structure as IM16, even when the two grammars derive essentially the same accent pattern on the surface. I attribute this "failure" of the model to general difficulties in the learning of so-called hidden structures such as metrical feet.

To conclude, this study offers a first attempt to apply IM16's formal analysis of Japanese accent to large-scale loanword data using a probabilistic grammar. Its overall success lends support to IM16. It also raises broader implications for the field of phonology.

References

Goldwater, S. and Johnson, M. 2003. Learning OT constraint rankings using a maximum entropy model. *Proceedings of the Stockholm Workshop on Variation within Optimality Theory*. 111-120.

Inoue, F. 1998. *Nihongo-Watching [Japanese-Watching]*. Iwanami Shoten: Tokyo.

Ito, J. and Mester, A., 2016. Unaccentedness in Japanese. *Linguistic Inquiry* 47:3. 471–526.

Mutsukawa, M., 2009. *Japanese Loanword Phonology: The Nature of Inputs and the Loanword Sublexicon*. Nanzan University Monograph Series. Hituzi Syobo: Tokyo.