Predicting semi-regular Japanese accent through gradient feature strengths

Eric Rosen, University of British Columbia

Motivation for feature strength Surface pitch accent of Japanese noun-noun compounds in which neither conjunct exceeds two moras is considered at best semi-predictable, thus requiring lexical listing of the accent locus of each compound. (Haruo Kubozono, personal communication, Kubozono and Fujiura 2004) This paper argues that these patterns can be predicted solely from input forms of constituents if input activations for an accent feature can have strengths that vary across lexical items but remain constant for a given item. In the framework of Gradient Symbolic Computation (Smolensky and Goldrick 2015, henceforth GSC), adopted here, inputs can have partial activations and constraints are weighted.

Minimal pairs suggest a strength hierarchy of accent features In (1) below, a subset of over 1000 such compounds, morphologically minimal pairs occur with the same N_1 in horizontal pairs and with the same N_2 along upwards-right pairs. We show that the accent can be determined by the combined strength of accent activation features on N_1 and N_2, and that a corpus of data exhibits a hierarchy of activation strength like that in (2) for N_1s and (3) for N_2s, where N_i accenting with N_j but not with N_k entails a strength difference $N_j \gg N_k$. Significantly, we never find compounds that contradict this hierarchy: e.g. no N_1 preaccents with *kawá* 'hide' but not with *mizu* 'water'.

Coalescence of floating accent features can explain combined effect of constituents on accentuation Given that the choice of an N_2 arguably affects accent occurrence on N_1 as in (1a,b) and N_1 affects N_2 as in *huyu-zora* 'winter sky' (unaccented) vs. *aki-zóra* 'autumn sky' (accent-keeping), we posit floating accent features at the edges of each noun that can coalesce in the output with a heteromorphemic accent feature with strengths contributing additively as shown in (4). We posit that a highly-ranked, strict LINEARITY constraint rules out coalescence of two tautomorphemic features.

A sample derivation The two harmonic tableaux in (5) and (6) account for accentuation in (1c) vs. non-accentuation in (1d) through proposed, differing input accent feature strengths (indicated with bracketed decimal numbers) on *así* 'foot' and *kuti* 'mouth'. In GSC, the harmony of a candidate is the sum of the harmonies produced by each constraint, with the most harmonic candidate being optimal. Satisfaction of a MAX constraint contributes positive harmony to the amount of a feature's input activation that surfaces. DEP constraints contribute negative harmony of the deficit between output and input activations. GSC's process of *quantization* forces output activations to be at or near 0 or 1, so we only consider candidates with those values, and here we consider quantization to operate globally, where at most one mora surfaces with accent. We assume initial and final accent to be ruled out by high-ranked constraints (not shown in tableaux) RIGHTMOSTFT (no Ft follows head Ft), NOLAPSE (violated by 2 consecutively unparsed σ's), NONFINAL(σ) (no accent on word-final σ) and MORAICTROCHEE, following the analysis of Itô and Mester (2016).

Learning An error-driven learning algorithm was able to find activation values for accent features in a database of N-N compounds that correctly derived accent patterns for both the set of compounds and also the simplex noun constituents given the same feature strengths.

Summary The uniformity of structural positions throughout the dataset of $2\mu + 2\mu$ compounds rules out the possibility that gradient underlying tendencies towards accentuation can be attributed to prosodic, morphological or contextual factors. Positing gradient feature strengths as a lexical property accounts for this set of morphologically complex surface patterns that are unpredictable if input features cannot have varying strengths.

Data

(1)

	Preaccented			Unaccented	
a.	*amá-mizu*	'rain-water'	b.	*ama-asi*	'rain-foot' (beating of the rain)
c.	*ató-asi*	'hind foot'	d.	*ato-kuti*	'after-mouth' (aftertaste)
e.	*waní-guti*	'crocodile-mouth' (folklore creature)	f.	*wani-gawa*	'crocodile hide'

(2) *wáni* 'crocodile' $\gg$ *áto* 'after' $\gg$ *áme* 'rain' (by comparing e. to d. and c. to b.)

(3) *mizu* 'water' $\gg$ *así* 'foot' $\gg$ *kuti* 'mouth' $\gg$ *kawá* 'hide'
(by comparing a. to b., c. to d. and e. to f.)

(4)

$$\text{Input}: \begin{array}{cccccccc} a_1 & a_2 & a_3 & \vdots & a_4 & a_5 & a_6 \\ | & | & & \vdots & & | & | \\ \mu_1 & \mu_2 & & \vdots & & \mu_3 & \mu_4 \end{array} \qquad \text{Output}: \begin{array}{cccc} a_1 & a_2 + a_4 & a_3 + a_5 & a_6 \\ | & | & | & | \\ \mu_1 & \mu_2 & \mu_3 & \mu_4 \end{array}$$

(5)

$(0.35)_1$ $(0.3)_2$ $(0)_3$ $(0.25)_4$ $(0)_5$ $(0.2)_6$ — a t o + a s i	Max_1	Max_2	Max_3	Max_4	Max_5	Max_6	Dep	H
(a) ☞ **ató-asi**		0.3		0.25			-0.45	**0.1**
(b) ato-ási							-1.0	-1.0
(c) ato-asi								0

(A separate MAX constraint calculation for each indexed feature.) For cand. (a), activation sum of coalescing a_2 and a_4, $(0.3 + 0.25 = 0.55)$ due to MAX exceeds deficit $(1.0 - 0.55 = 0.45)$ due to DEP, giving net Harmony 0.1. For cand. (b), no input activations for accenting initial mora of *asi* result in zero Harmony due to MAX.

(6)

$(0.35)_1$ $(0.3)_2$ $(0)_3$ $(0.15)_4$ $(0)_5$ $(0)_6$ — a t o + k u t i	Max_1	Max_2	Max_3	Max_4	Max_5	Max_6	Dep	H
(a) ató-kuti		0.3		0.15			-0.55	-0.1
(b) ato-kúti							-1.0	-1.0
(c) ☞ **ato-kuti**								**0**

Lower activation on floating a_4 of *kuti* (0.15) than on *asi* (0.25) results in negative Harmony for preaccented candidate (a), making unaccented candidate (c) optimal.

References

Itô, Junko and Armin Mester. 2016. Unaccentedness in Japanese. *Linguistic Inquiry* 47. 471-526.

Kubozono, Haruo and Yayoi Fujiura. 2004. Morpheme-dependent nature of compound accent in Japanese: An analysis of "short" compounds. *On-in Kenkyuu* [phonological studies] 7. 916.

Smolensky, Paul and Matthew Goldrick. 2015. Gradient Symbolic Computation. LSA Summer Institute Workshop. Chicago.