Evidence for gradient input features from Sino-Japanese compound accent

Eric Rosen, Johns Hopkins University, errosen@mail.ubc.ca

Predicting the pitch-accent patterns of two-member Sino-Japanese compounds presents an analytic challenge (Kawahara 2015:17). Although both morphemes (henceforth N_1 and N_2), show general accenting tendencies, we show, using a corpus of compounds from the NHK Accent Dictionary, that deriving their accent patterns without lexically listing each compound accent requires gradient feature values, as in the Gradient Symbolic Computation framework (Smolensky and Goldrick 2015, henceforth GSC). Both prosody and morphological identity affect accent in 2-morpheme structures but even with morpheme-specific constraints, both Optimality Theory (Prince and Smolensky 1993) and Harmonic Grammar (Pater 2009) fail to provide an explanatory analysis. The problem is that the tendency of a morpheme to trigger accent on its neighbour operates differently left-to-right than right-to-left, as shown by the contrast between 社会 *syá-kai* (accented) 'society' (lit. company-meet) and 会社 *kai-sya* (unaccented) 'company', with morpheme order switched. Prosody alone cannot explain this contrast, given the abundance of pairs like *ha-tyoo* (unaccented) 波長 'wave-length' and *tyóo-ha* 長波 'long-wave' (accented) with the opposite accent-prosody correlation.

A GSC analysis succeeds because the formalism naturally affords three locations for accent-affecting propensity: an underlyingly mora-associated position and floating positions on the left and right edges. A simple machine-learning algorithm finds *accent-affecting propensities = activations* that collectively work for a set of compounds with frequently-occurring morphemes from the NHK corpus.

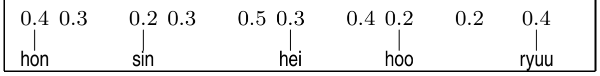
Each single-character morpheme is $\leq 2\mu$, e.g. *kái-gai* 海外 'overseas' (lit. ocean-outside). Morphemes show gradient accenting tendencies: e.g., *dai* 代 'era', is accent-friendly, triggering accent as N_2 in 13/18 compounds: e.g. *kó-dai* 古代 'ancient times', but fails to trigger in *nen-dai* 年代 'generation, age'. In contrast, *sei/syoo* 性 'nature', blocks accent in 13/14 compounds: e.g. *tyuu-sei* 中性 'neutral', but triggers accent in *tén-sei* 天性 'second-nature'.

We show that accent is determined both by prosody (Itô and Mester 2016, henceforth I&M) and combined accenting tendencies of N_1 and N_2. In the prosodically identical and morphologically minimal pairs in (1), with contrasting accentuation shown by shading, accent cannot be determined by N_1 alone or N_2 alone: *hon* 'main', *hoo* 'law' and *sin* 'new' all variably affect accenting.

(1) h → p by rule	(a) *hón-poo* 本法 'this-law'	(b) *sin-poo* 新法 'new-law'
	(c) *hon-ryuu* 本流 'main-stream'	(d) *sín-pei* 新兵 'new-recruit'

We posit underlying accent features with gradient activation that are anchored to moras or float at the left and/or right morpheme edge:

(2) Sample underlying accent activations of morphemes



A mora-linked accent feature can coalesce with a floating feature on the adjacent morpheme: e.g. 0.4 on *hon* with floating 0.4 on *hoo*, resulting in 0.4+0.4=0.8. Only if this additive activation exceeds some threshold, determined by weighted MAX and DEP constraints, will an accent surface. As shown in (2)-(5) below, GSC MAX constraints contribute positive Harmony to a candidate to the extent to which its underlying activation surfaces (e.g. 0.8 in *hón-poo*); DEP constraints negative Harmony for the deficit between an underlying value and full activation in a candidate (e.g. 1-0.8=0.2 in *hón-poo*). The three prosodic constraints, RIGHTMOST, INITIALFT, WDACCENT are adapted from I&M but with weighted values rather than categorical ranking. The winning candidate (bolded) has the highest Harmony value.

(3) hon+hoo	MAX +1	DEP -1	RMOST -0.5	INFT -0.8	WDACC -0.25	H
hón-poo 0.4 + 0.4	0.8	-0.2	-0.5			0.1
hon-póo 0.3 +0.2	0.5	-0.5		-0.8		-0.8
hon-poo					-0.25	-0.25

(4) hon+ryuu	Max +1	Dep -1	Rmost -0.5	InFt -0.8	WdAcc -0.25	H
hón-ryuu 0.4 + 0.2	0.6	-0.4	-0.5			-0.3
hon-rýuu 0.3 + 0.4	0.7	-0.3		-0.8		-0.4
hon-ryuu					-0.25	-0.25

(5) sin+hoo	Max +1	Dep -1	Rmost -0.5	InFt -0.8	WdAcc -0.25	H
sín-poo 0.2 + 0.4	0.6	-0.4	-0.5			-0.3
sin-póo 0.3 + 0.2	0.5	-0.5		-0.8		-0.8
sin-poo					-0.25	-0.25

(6) sin+hei	Max +1	Dep -1	Rmost -0.5	InFt -0.8	WdAcc -0.25	H
sín-pei 0.2 + 0.5	0.7	-0.3	-0.5			-0.1
sin-péi 0.3 + 0.3	0.6	-0.4		-0.8		-0.6
sin-pei					-0.25	-0.25

Standard OT or HG with morpheme-specific constraints cannot handle pairs like *syá-kai* 'society' and *kai-sya* 'company' above. Because morpheme *kai* **can** bear accent in other $\mu\mu$-μ compounds such as 会期 *kái-ki* 'legislative session', OT or HG needs to measure the combined action of N$_1$ and N$_2$ as in the present account, e.g. through coalescence of features, allowing indexed constraints from the two coalescing morphemes to act together. *kai-sya* and *káiki* can only be derived with a ranking (7) that will be contradictory for *syá-kai*. (α= accent feature)

(7) $\alpha_1\,\alpha_2$ $\alpha_3\alpha_4$ $\alpha_5\,\alpha_6$ kai ki sya	*Acc sya	MaxAcc kai	*Acc ki
kái-ki ($\alpha_1 + \alpha_3$) kai-ki		*!	*
kái-sya ($\alpha_1 + \alpha_5$) **kai-sya**	*!	*	

This ranking, necessary for (7), fails for *syá-kai* in (8), since the constraints, whether weighted or categorical, will operate the same way as in unaccented *kai-sya* for both OT and HG.

(8) $\alpha_1\,\alpha_2$ $\alpha_3\,\alpha_4$ sya kai	*Acc sya	MaxAcc kai	*Acc ki
syá-kai ($\alpha_1 + \alpha_3$) ☹sya-kai	*!	*	

GSC, which allows gradient activations, is able to derive these two contrasting compounds, since different activations can occur on different accent features: floating at L and/or R edges (shown in (9)-(11) by *L* and *R* subscripts) and μ-anchored (*A* subscript) :

Some possible input accent activations for (9)-(11)		
$.2_L$ $.3_A$ $.2_R$ sya	$.4_L$ $.2_A$ $.2_R$ kai	$.5_L$ $.2_A$ $.2_R$ ki

(9) sya+kai	Max +1	Dep -1	Rmost -0.5	InFt -0.8	WdAcc -0.25	H
☞ syá-kai $0.3_A + 0.4_L$	0.7	-0.3	-0.5			-0.1
sya-kái $0.2_R + 0.2_A$	0.4	-0.6		-0.8		-1.0
sya-kai					-0.25	-0.25

(10) kai+sya	Max +1	Dep -1	Rmost -0.5	InFt -0.8	WdAcc -0.25	H
kái-sya $0.2_A + 0.2_L$	0.4	-0.6	-0.5			-0.7
kai-syá $0.2_R + 0.3_A$	0.5	-0.5		-0.8		-0.8
☞ kai-sya					-0.25	-0.25

(11) kai+ki	Max +1	Dep -1	Rmost -0.5	InFt -0.8	WdAcc -0.25	H
☞ kái-ki $.2_A + .5_L$	0.7	-0.3	-0.5			-0.1
kai-kí $.2_R + .2_A$	0.4	-0.6		-0.8		-1.0
kai-ki					-0.25	-0.25

This derivation is not possible when constraints are simply indexed to morphemes, unless a linked feature and a floating feature for a stem were, implausibly, separate morphemes with different constraint indices. The gradient-activation approach also moves the semi-regularity of gradient accenting tendencies of morphemes from the grammar to its arguably proper place in the lexicon.

To test learnability, a simulation using logistic regression, a well-tested method for classification, was applied to the set of 663 compounds in the corpus with unambiguous accent in which both morphemes occurred at least 50 times, and was able to find input values for accentuation that correctly derived their accent patterns in the GSC framework.

References

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

Kawahara, Shigeto. 2015. The phonology of Japanese accent. In *Handbook of Japanese phonetics and phonology*, Haruo Kubozono, ed. De Gruyter Mouton. Berlin; Boston.

Pater, Joe. 2009. Weighted Constraints in Generative Linguistics. *Cognitive Science* 33:999–1035.

Prince, Alan and Paul Smolensky. 1993. Optimality Theory. Ms. Rutgers University.

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