



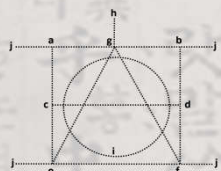
Korean orthography and phonology: What's the correlation?

Hana Jee, Monica Tamariz, and Richard Shillcock

Hangeul, Korean orthography is well known for its letter-sound connections. But how strong are they? We quantified this connection and suggested a method allowing comparison with other orthographies.

Distance between letters?

Stroke Share Rate



IPA	a-b	c-d	e-f	g-h	i-j	k-l	m-n	o-p	q-r	s-t	u-v	w-x	y-z	digit
ㄱ	/g/	1	0	0	1	1	0	0	0	0	0	0	0	0
ㅋ	/k/	1	1	0	1	1	0	0	0	0	0	0	0	0
ㆁ	/ŋ/	1	0	0	1	1	0	0	0	0	0	0	0	1
*	/h/	1	0	0	0	0	0	0	0	1	1	0	0	0

- Each letter defined as a binary vector.
- Distance between two vectors
 - Euclidean distance
 - Feature edit distance

Hausdorff Distance*

- Implemented in Python 3.6
- Compare images pixel by pixel



Distance between phonemes?

International Phonetic Alphabet

- Feature based phoneme description
- Each letter defined as a binary vector.
- Distance between two vectors
 - Euclidean distance
 - Feature edit distance

	Place of Articulation						Manner of Articulation					
	Hard palatal	Soft palatal	Labial	Dental	Throat	Plosive	Affricative	Lenis	Aspirated	Tensed	Nasal	Flow
ㄱ	1	0	0	0	0	1	0	1	0	0	0	0
ㅋ	1	0	0	0	0	1	0	0	1	0	0	0
ㆁ	1	0	0	0	0	1	0	0	0	1	0	0

Finding 1. Medium to high correlation Result 1
Finding 2. Euclidean distance better Result 1
Finding 3. Stronger in sans-serif fonts Result 2

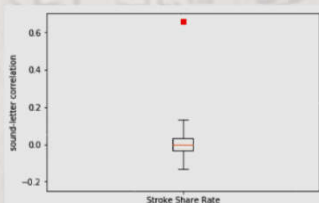
Result 1

Measurement	Correlation coefficient (rho)	p-value
Euclidean	0.66	<2.2e-16 ***
Feature edit	0.53	<2.2e-16 ***

- Medium to high letter-sound correlation
- Stronger correlation with Euclidean distance

	rho	p-value
Consonant	0.26	0.01 **
Vowels	0.49	0.001 ***

- Partial correlation test
- Larger contribution of vowels to systematicity



- Monte-Carlo permutation test confirms the correlation was not by chance

Result 2

	rho	p-value
나눔고딕	0.365	0.000 ***
맑은고딕	0.315	0.000 ***
나눔바른펜	0.314	0.000 ***
나눔바른고딕	0.304	0.000 ***
Noto Serif CJK KR	0.303	0.000 ***
나눔명조	0.300	0.000 ***
굴림	0.299	0.000 ***
궁서	0.277	0.000 ***
조선일보명조체	0.269	0.000 ***
EBS주시정체	0.240	0.000 ***
tvN출거위키키체	0.233	0.000 ***
한겨레결체	0.229	0.000 ***
돋움	0.188	0.000 ***
바탕	0.183	0.000 ***
KCC 길훈	0.182	0.000 ***
신비논일곱살	0.159	0.001 ***
나눔손글씨박체	0.105	0.034 *
신라자	0.104	0.035 *

- Stronger sound letter correlation in some fonts
- Sans-serif fonts work better than others.
- Hausdorff distance: possible to compare with other orthographies.

*Huttenlocher, D. P., Rucklidge, W. J., & Klanderman, G. A. (1992, June). Comparing images using the Hausdorff distance under translation. In Proceedings 1992 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (pp. 654-656). IEEE. **By Bibliography** [1] Monaghan, P., Shillcock, R. C., Christiansen, M. H., & Kirby, S. (2014). How arbitrary is language?. Philosophical Transactions of the Royal Society B: Biological Sciences, 369(1651), 20130299. [2] Tamariz, M. (2008). Exploring systematicity between phonological and context-cooccurrence representations of the mental lexicon. The Mental Lexicon, 3(2), 259-278.