Asian Culture and History; Vol. 15, No. 1; 2023
ISSN 1916-9655 E-ISSN 1916-9663
Published by Canadian Center of Science and Education

Quantified Grapho-Phonemic Systematicity in Korean Hangeul

Hana Jee[1], Monica Tamariz[2] & Richard Shillcock[1]

[1] Psychology, The University of Edinburgh, UK

[2] Psychology, Heriot-Watt University, UK

Correspondence: Hana Jee, Korean and Applied Linguistics, School of Education, Language and Psychology, York St John University YO31 7EX, UK.

Received: January 6, 2023 Accepted: January 26, 2023 Online Published: February 2, 2023

doi:10.5539/ach.v15n1p25 URL: https://doi.org/10.5539/ach.v15n1p25

Abstract

Hangeul, the Korean orthography is well known for its scientific design that emphasizes the link between sounds and letter shapes. However, it hasn't been asked so far 'how systematic' it is. We quantify, for the first time, the grapho-phonemic systematicity of hangeul. We defined Korean phonemes as binary vectors according to articulatory features and then measured the pairwise phonemic distance between phonemes using multiple methods. We measured the pairwise visual distance between letter shapes by (a) stroke share rate, which reflects the original principles of hangeul's creation, and (b) Hausdorff distance (Huttenlocher et al., 1993), which measures topological difference between images. We then tested the correlation between the phonological distances and the corresponding orthographical distances. Positive correlations clearly indicated that similar letters tend to have similar pronunciations in Korean hangeul. Stroke share rate maximizes hangeul's grapho-phonemic systematicity. Hausdorff distance, an initial step in the detailed quantifying of visual distance, allows similar calculations to be carried out with any hangeul font and with any other orthography (Jee, Tamariz, & Shillcock, 2021; 2022a; 2022b). Consciously designed to be phonologically transparent, hangeul can be considered as the gold standard of grapho-phonemic systematicity. We discuss the implications of this systematicity.

Keywords: hangeul, systematicity, Korean phonology, grapho-phonemic systematicity, Hausdorff distance

1. Introduction

Hangeul, the Korean orthography, is renowned for the availability of information about its origins. It is the only orthography that a king himself designed for the illiterate among his people. Named *Hunmin Jeongeum, the Standard Sounds for the Instruction of the People*, 28 letters were created in 1443 and promulgated in 1446. Hangeul has been highly appreciated by linguists and others worldwide. It has been dubbed 'the most scientific system of writing' (Reischauer & Fairbank, cited in Hyun, 1981) and 'unquestionably one of the great intellectual achievements of humankind (Sampson, 1985). The reasons for these commendations are: (i) it is orthographically shallow; (ii) it expresses fine phonemic distinctions; (iii) its letter shapes visualize articulation; and (iv) its letter shapes are consistent with the corresponding phonemes.

Orthographic depth is defined by the extent to which the letter-sound association is transparent and predictable (Seymour et al., 2003). Shallow orthography facilitates reading (Martin et al., 2016; Paulesu et al., 2001; Spencer, 2001). It is less effortful (Paulesu et al., 2000). Dyslexics experience fewer difficulties when reading shallow orthography than reading deep orthography (Paulesu et al., 2001). Hangeul is a shallow orthography, along with Finnish, Italian, and Turkish.

The smallest unit of sound specified by an orthography differs across languages. Chinese is based on syllables (consonant + vowel + selective final consonant): for example, 看 /kàn/; and 是 /shǐ. Japanese consists of *morae* (consonant + vowel): ご /go/; and ち /t͡ɕi/. English has an alphabet system, in which a letter is linked to a phoneme (either consonant or vowel). Korean orthography specifies phonological features (Sampson, 1985). For example, tensed phonemes are distinguished from tenseless phonemes: /t/ - /t͈/; and /k/ - /k͈/. With this fine phonological distinction, hangeul was designed to represent the sound of cranes, chickens, and even the wind (Hideki, 2011).

Along with the tensed phonemes, Korean consonantal sounds consist of sets of three: *lenis*, *aspirated* and *tensed*. In hangeul, these phoneme sets are represented as visually consistent letters. In general, adding a stroke makes the lenis letter aspirated (e.g., ㄱ /g/ - ㅋ /k/) and duplicating the letter makes it tensed (e.g., ㄱ /g/ -ㄲ / k̬/). Overall, a set of the phonemes that share an articulation point have visually similar letter-shapes. Likewise, diphthongs are distinguished from monophthongs by an additional stroke (e.g., ㅏ/a/ - ㅑ /ja/).

Finally, Korean consonants attempt iconically to visualize the shape of articulation. The letter ㄱ /g/ represents the tongue touching the hard palate (seen from the left); ㅅ /s/ represents the airflow through the teeth; and ㅇ /ŋ/ represents the throat and does not have phonemic value at the onset. Not all consonants are pictographic. ㅁ /m/, presumably came from Chinese 口 /kǒu/, meaning 'mouth' (Sampson, 1985). The obstruents (ㄷ /d/, ㅈ /tɕ/, and ㅂ /b/) are distinguished from the continuants (ㄴ /n/, ㅅ /s/, and ㅁ /m/) by additional strokes. Meanwhile, Korean vowels have a cultural basis. They are composed by the combination of the earth (ㅡ), the heaven (ㆍ), and human being (ㅣ), symbolizing harmony among all three.

Given the unique compositionality of hangeul, there has been little or no attempt to quantify the relation between letters and phonemes. Can this be done? Is letter-sound systematicity in hangeul indeed greater than those in other orthographies?

2. Procedure

Our grapho-phonemic analysis follows the principles of phono-semantic research (Dautriche et al., 2017; Monaghan et al., 2014; Tamariz, 2008). We measured all possible pairwise visual distances between letter shapes and all the corresponding pairwise phonological distances between phonemes. Then we measured the correlation between these two long lists of corresponding distances. Finally, to verify its statistical significance, we conducted a Monte-Carlo permutation test, as in the literature on sound-meaning systematicity.

We expected hangeul to return a robust, positive correlation between orthographical distances and phonological distances, considering the principles of its creation. A positive correlation indicates that similar letters tend to have similar sounds. We also expected that the level of systematicity in hangeul would be higher than in other less insightfully created orthographies.

2.1 Measuring Phonological Distance

We studied 19 consonants (ㄱ, ㄴ, ㄷ, ㄹ, ㅁ, ㅂ, ㅅ, ㅇ, ㅈ, ㅊ, ㅋ, ㅌ, ㅍ, ㅎ, ㄲ, ㄸ, ㅃ, ㅆ, ㅉ) and 10 monophthongs (ㅏ, ㅓ, ㅗ, ㅜ, ㅐ, ㅔ, ㅚ, ㅟ, ㅡ, ㅣ). Each phoneme was defined as a binary vector according to its articulatory features (Table 1).

Table 1. Articulatory features of Korean phonemes

Letter	Phoneme	Place of Articulation					Manner of Articulation								Vowel Quality						
		Hard palatal	Soft palatal	Labial	Dental	Throat	Plosive	Affricate	Fricative	Fortis	Lenis	Aspirated	Nasal	Flow	Front	Back	High	Middle	Low	Round	Non-round
ㅋ	k^h	0	1	0	0	0	1	0	0	0	0	1	0	0	0	0	0	0	0	0	0
ㄱ	g	0	1	0	0	0	1	0	0	0	1	0	0	0	0	0	0	0	0	0	0
ㄲ	k*	0	1	0	0	0	1	0	0	1	0	0	0	0	0	0	0	0	0	0	0
ㄴ	n	0	0	0	1	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0
ㅌ	t^h	0	0	0	1	0	1	0	0	0	0	1	0	0	0	0	0	0	0	0	0
ㄷ	d	0	0	0	1	0	1	0	0	0	1	0	0	0	0	0	0	0	0	0	0
ㄸ	t*	0	0	0	1	0	1	0	0	1	0	0	0	0	0	0	0	0	0	0	0
ㄹ	l	0	0	0	1	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0
ㅁ	m	0	0	1	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0
ㅍ	p^h	0	0	1	0	0	1	0	0	0	0	1	0	0	0	0	0	0	0	0	0
ㅂ	b	0	0	1	0	0	1	0	0	0	1	0	0	0	0	0	0	0	0	0	0
ㅃ	p̬*	0	0	1	0	0	1	0	0	1	0	0	0	0	0	0	0	0	0	0	0
ㅅ	s	0	0	0	1	0	0	0	1	0	1	0	0	0	0	0	0	0	0	0	0
ㅆ	s*	0	0	0	1	0	0	0	1	1	0	0	0	0	0	0	0	0	0	0	0
ㅇ	ŋ	0	1	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0
ㅊ	$tɕ^h$	1	0	0	0	0	0	1	0	0	0	1	0	0	0	0	0	0	0	0	0
ㅈ	tɕ	1	0	0	0	0	0	1	0	0	1	0	0	0	0	0	0	0	0	0	0
ㅉ	t̬ɕ*	1	0	0	0	0	0	1	0	1	0	0	0	0	0	0	0	0	0	0	0
ㅎ	h	0	0	0	0	1	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0
ㅏ	a	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	1	0	1

ㅐ	ɛ	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	1	0	1
ㅓ	ʌ	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	1	0	0	1
ㅔ	e	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	1	0	0	1
ㅗ	o	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	1	0	1	0
ㅚ	ɸ	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	1	0	1	0
ㅜ	u	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	0	0	1	0
ㅟ	y	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	1	0	0	1	0
ㅡ	ɯ	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	0	0	0	1
ㅣ	i	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	1	0	0	0	1

The phonological distance between two phonemes was defined as the distance between their vectors (Monaghan et al., 2010). We used different distance measures to ensure robustness: Feature edit distance, the number of different features between two vectors; Euclidean distance, the shortest geometric distance between two vectors; Cosine distance, the angle made by the two vectors; and Jaccard distance, the number of shared features divided by the total number of features. The first two were used by Monaghan et al. (2014).

2.2 Measuring Orthographical Distance

2.2.1 Stroke Share Rate

Comparing salient sub-letter features to measure the visual difference between letters is not a new idea (Briggs & Hocevar, 1975; Geyer & DeWald, 1973; Watt, 1979). However, there is little or no research on hangeul letters from this perspective. We designed a novel method specifically for hangeul. First, we decomposed the letters into strokes and defined them topologically (Figure 1). We then re-defined each letter as a binary vector (Table 2). Thus, the distance between two letters now equals the distance between two 19-place vectors (12 places for the consonants, 7 for the vowels).

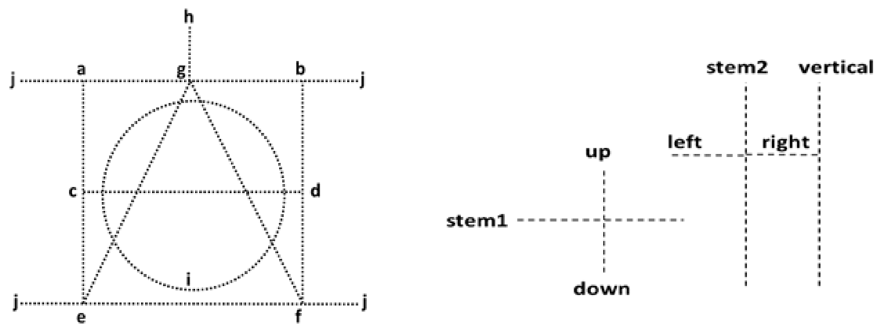

Figure 1. Decomposition of hangeul consonants (left) and vowels (right) into strokes

Table 2. Orthographic features of Korean letters for stroke-share rate

		Consonants												Vowels						
Letter	IPA	a-b	c-d	e-f	b-d	d-f	a-c	c-e	j	i	g-h	g-e & g-f	duplicated	stem1	up	down	stem2	right	left	vertical
ㅋ	k^h	1	1	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
ㄱ	g	1	0	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
ㄲ	k*	1	0	0	1	1	0	0	0	0	0	0	1	0	0	0	0	0	0	0
ㄴ	n	0	0	1	0	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0
ㅌ	t^h	1	1	1	0	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0
ㄷ	d	1	0	1	0	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0
ㄸ	t*	1	0	1	0	0	1	1	0	0	0	0	1	0	0	0	0	0	0	0
ㄹ	l	1	1	1	1	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0
ㅁ	m	1	0	1	1	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0
ㅍ	p^h	1	0	1	1	1	1	1	1	0	0	0	0	0	0	0	0	0	0	0
ㅂ	b	0	1	1	1	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0
ㅃ	p*	0	1	1	1	1	1	1	0	0	0	0	1	0	0	0	0	0	0	0
ㅅ	s	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0

ㅆ	s*	0	0	0	0	0	0	0	0	0	0	1	1	0	0	0	0	0	0	0
ㅇ	ŋ	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0
ㅊ	tɕ^h	1	0	0	0	0	0	0	0	0	1	1	0	0	0	0	0	0	0	0
ㅈ	tɕ	1	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0
ㅉ	t͈ɕ*	1	0	0	0	0	0	0	0	0	0	1	1	0	0	0	0	0	0	0
ㅎ	h	1	0	0	0	0	0	0	0	1	1	0	0	0	0	0	0	0	0	0
ㅏ	a	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	0	0
ㅐ	ɛ	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	0	1
ㅓ	ʌ	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	1	0
ㅔ	e	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	1	1
ㅗ	o	0	0	0	0	0	0	0	0	0	0	0	0	1	1	0	0	0	0	0
ㅚ	ϕ	0	0	0	0	0	0	0	0	0	0	0	0	1	1	0	1	0	0	0
ㅜ	u	0	0	0	0	0	0	0	0	0	0	0	0	1	0	1	0	0	0	0
ㅟ	y	0	0	0	0	0	0	0	0	0	0	0	0	1	0	1	1	0	0	0
ㅡ	ɯ	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0
ㅣ	i	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0

As with phonological distances, the orthographic distances were measured by four different metrics: feature edit distance, Euclidean distance, Cosine distance, and Jaccard distance.

2.2.2 Hausdorff Distance

We present stroke share rate, above, as a point of comparison with the detailed quantitative measure of Hausdorff distance (Huttenlocher et al., 1993). Unlike the hangeul-specific stroke share rate, Hausdorff distance can be applied to any script system because it treats the letters as images. It converts the image into a black and white raster graphic. Given two sets of black pixels, $X = \{x_1, \dots x_n\}$ and $Y = \{y_1, \dots y_n\}$, the directed Hausdorff distance is calculated as follows:

$$d_h(X,Y) = max(\underset{x \in X}{max}, min_{y \in Y}\ |x - y|) \tag{1}$$

where Euclidean distance measures the distance between two individual points, $|x - y|$. Being fundamentally asymmetric ($d(X, Y) \neq d(Y, X)$), the larger value between the two (max) is returned. Because Hausdorff distances recognizes letters as images, different fonts return different values. We examined 88 available Korean fonts. Scipy.spatial. distance.directed_hausdorff (ver. 1.3.1) was implemented on Python 3.6.1 (Note 1).

3. Results

We calculated the correlation between the two lists of corresponding visual and phonological distances using Pearson's *r*, to quantify grapho-phonemic systematicity. The results were separately presented below according to the orthographic distance measure.

3.1 Grapho-Phonemic Systematicity (Stroke Share Rate)

Table 3 shows the correlation coefficients (Pearson's *r*) between the orthographic distances and the phonological distances, when stroke share rate was used to measure the orthographic distances.

Positive correlation coefficients in general mean that similar letter shapes tend to have similar sounds, or vice versa, which quantitatively confirms the principle based on which hangeul was created. Very low p-values indicate the significance of the statistical analysis.

Table 3. Grapho-phonemic systematicity of hangeul when orthographic distances were measured by stroke share rate

Phonological distance measure	*r*	*p*
Euclidean	0.51	< .001
Cosine	0.60	< .001
Jaccard	0.60	< .001
Feature edit	0.51	< .001

Note. * $p < .05$, ** $p < .01$, *** $p < .001$.

3.2 Grapho-Phonemic Systematicity (Hausdorff Distance)

The orthographic distances were also measured by Hausdorff distance. Table 4 shows the grapho-phonemic systematicity from 88 Korean fonts. The majority displayed very significant correlations between letters and sounds although the correlation coefficients are not as high as those in Table 3. The results are robust across the phonological distance measures.

Table 4. Grapho-phonemic systematicity of 88 fonts when orthographic distances were measured by Hausdorff distance

	Euclidean			Cosine			Jaccard			feature edit		
Font	*r*	*p*		*r*	*p*		*r*	*p*		*r*	*p*	
굴림	0.16	0.00	***	0.11	0.02	*	0.11	0.03	*	0.24	0.00	***
돋움	0.05	0.28		0.03	0.52		0.03	0.52		0.12	0.02	*
바탕	0.09	0.07	.	0.07	0.14		0.06	0.20		0.18	0.00	***
궁서	0.22	0.00	***	0.18	0.00	***	0.16	0.00	***	0.30	0.00	***
맑은고딕	0.10	0.05	*	0.08	0.12		0.07	0.15		0.18	0.00	***
나눔고딕	0.11	0.03	*	0.08	0.12		0.07	0.14		0.18	0.00	***
나눔명조	0.12	0.02	*	0.07	0.14		0.07	0.15		0.20	0.00	***
나눔손글씨붓체	0.24	0.00	***	0.22	0.00	***	0.21	0.00	***	0.23	0.00	***
나눔손글씨 펜체	0.18	0.00	***	0.16	0.00	***	0.16	0.00	***	0.15	0.00	**
나눔바른고딕	0.11	0.03	*	0.08	0.12		0.08	0.11		0.18	0.00	***
나눔바른펜	0.16	0.00	**	0.12	0.02	*	0.10	0.04	*	0.27	0.00	***
나눔스퀘어	0.09	0.06	.	0.06	0.21		0.06	0.20		0.16	0.00	***
나눔스퀘어라운드	0.10	0.05	.	0.06	0.20		0.07	0.18		0.16	0.00	***
Noto Sans CJK KR	0.10	0.04	*	0.08	0.12		0.08	0.10		0.16	0.00	***
NotoSerif CJK KR	0.10	0.05	.	0.06	0.21		0.06	0.24		0.18	0.00	***
도현체	0.13	0.01	**	0.10	0.05	*	0.09	0.06	.	0.23	0.00	***
주아체	0.12	0.02	*	0.09	0.08	.	0.08	0.09	.	0.20	0.00	***
한나는 11 살체	0.19	0.00	***	0.16	0.00	***	0.15	0.00	**	0.29	0.00	***
간이벽온방	0.25	0.00	***	0.22	0.00	***	0.21	0.00	***	0.34	0.00	***
대한민국독도체	0.16	0.00	***	0.14	0.01	**	0.12	0.01	*	0.21	0.00	***
법정체	0.13	0.01	**	0.09	0.08	.	0.08	0.12		0.21	0.00	***
대한체	0.16	0.00	***	0.11	0.02	*	0.11	0.03	*	0.24	0.00	***
월인석보체	0.14	0.01	**	0.12	0.02	*	0.11	0.02	*	0.22	0.00	***
고도체	0.17	0.00	***	0.14	0.01	**	0.14	0.01	**	0.23	0.00	***
아리따돋움	0.18	0.00	***	0.16	0.00	***	0.15	0.00	**	0.28	0.00	***
HS 봄바람체 2.0	0.24	0.00	***	0.19	0.00	***	0.19	0.00	***	0.31	0.00	***
HS 가을생각체	0.26	0.00	***	0.21	0.00	***	0.20	0.00	***	0.37	0.00	***
HS 겨울눈꽃체	0.18	0.00	***	0.13	0.01	**	0.13	0.01	*	0.23	0.00	***
HS 두꺼비체	0.24	0.00	***	0.18	0.00	***	0.18	0.00	***	0.35	0.00	***
가비아솔미체	0.13	0.01	**	0.10	0.05	.	0.09	0.07	.	0.21	0.00	***
가비아납작블럭체	0.13	0.01	*	0.09	0.08	.	0.08	0.09	.	0.20	0.00	***
미생체	0.15	0.00	**	0.12	0.01	*	0.12	0.02	*	0.18	0.00	***
신비는일곱살	0.05	0.33		0.02	0.62		0.03	0.52		0.06	0.26	
꽃길	0.25	0.00	***	0.20	0.00	***	0.20	0.00	***	0.25	0.00	***
개미똥구멍	0.15	0.00	**	0.10	0.05	*	0.09	0.06	.	0.17	0.00	***
신과장	0.21	0.00	***	0.16	0.00	**	0.15	0.00	**	0.23	0.00	***
제주한라산체	0.19	0.00	***	0.17	0.00	***	0.16	0.00	***	0.15	0.00	**
제주고딕체	0.08	0.11		0.05	0.33		0.05	0.29		0.14	0.00	**
제주명조체	0.19	0.00	***	0.14	0.00	**	0.14	0.01	**	0.28	0.00	***
부산체	0.09	0.07	.	0.07	0.14		0.06	0.20		0.18	0.00	***
고양체	0.13	0.01	**	0.09	0.07	.	0.09	0.07	.	0.17	0.00	***
고양일산체	0.13	0.01	*	0.10	0.05	*	0.10	0.05	.	0.19	0.00	***
오성과한음체	0.10	0.05	.	0.06	0.22		0.06	0.22		0.16	0.00	***
막걸리체	0.16	0.00	***	0.13	0.01	**	0.13	0.01	**	0.18	0.00	***
전라북도체	0.11	0.02	*	0.10	0.05	.	0.09	0.06	.	0.17	0.00	***

푸른전남체	0.09	0.07	.	0.05	0.29		0.05	0.29		0.16	0.00	***
KoPub 돋움체	0.05	0.28		0.03	0.58		0.03	0.54		0.12	0.02	*
KoPub 바탕체	0.12	0.02	*	0.08	0.11		0.08	0.12		0.21	0.00	***
다온청년고딕	0.07	0.14		0.04	0.38		0.05	0.35		0.15	0.00	**
EBS 주시경체	0.07	0.15		0.06	0.22		0.06	0.26		0.14	0.00	**
EBS 훈민정음	0.09	0.07	.	0.09	0.07	.	0.10	0.05	*	0.12	0.01	*
EBS 훈민정음새론체	0.13	0.01	**	0.12	0.02	*	0.12	0.02	*	0.20	0.00	***
KBIZ 한마음고딕	0.24	0.00	***	0.19	0.00	***	0.18	0.00	***	0.37	0.00	***
KBIZ 한마음명조	0.23	0.00	***	0.17	0.00	***	0.16	0.00	***	0.35	0.00	***
도서관체	0.21	0.00	***	0.15	0.00	**	0.14	0.00	**	0.33	0.00	***
호국체	0.17	0.00	***	0.16	0.00	***	0.16	0.00	***	0.27	0.00	***
이롭게바탕체	0.16	0.00	***	0.12	0.01	*	0.12	0.02	*	0.25	0.00	***
tvN 즐거운이야기체	0.17	0.00	***	0.14	0.01	**	0.13	0.01	**	0.22	0.00	***
티몬몬소리체	0.14	0.00	**	0.12	0.02	*	0.12	0.02	*	0.23	0.00	***
빙그레체	0.19	0.00	***	0.15	0.00	**	0.15	0.00	**	0.28	0.00	***
스웨거체	0.17	0.00	***	0.13	0.01	**	0.12	0.02	*	0.28	0.00	***
한겨레결체	0.02	0.72		0.01	0.88		0.00	0.99		0.08	0.13	
조선일보명조체	0.03	0.57		0.01	0.78		0.01	0.86		0.08	0.09	.
동그라미재단	0.15	0.00	**	0.12	0.02	*	0.11	0.02	*	0.22	0.00	***
FB 이철수 80 목판 TM	0.20	0.00	***	0.13	0.01	*	0.12	0.02	*	0.19	0.00	***
FB 이철수 80 목판 M	0.20	0.00	***	0.13	0.01	*	0.12	0.02	*	0.19	0.00	***
FB 이철수 90 목판 TM	0.18	0.00	***	0.15	0.00	**	0.14	0.00	**	0.15	0.00	**
FB 이철수 90 목판 M	0.18	0.00	***	0.15	0.00	**	0.14	0.00	**	0.15	0.00	**
FB 이철수 2000 목판 TM	0.23	0.00	***	0.19	0.00	***	0.19	0.00	***	0.24	0.00	***
FB 이철수 2001 목판 M	0.12	0.02	*	0.13	0.01	**	0.13	0.01	**	0.08	0.11	
FB 이철수 2001 목판 TM	0.12	0.02	*	0.13	0.01	**	0.13	0.01	**	0.08	0.11	
Yoon 다정	0.05	0.29		0.03	0.60		0.03	0.52		0.10	0.05	*
Yoon 민준	0.06	0.23		0.03	0.51		0.04	0.45		0.11	0.03	*
Yoon 세희	0.06	0.26		0.03	0.55		0.03	0.49		0.10	0.04	*
Yoon 아혜	0.09	0.06	.	0.06	0.26		0.06	0.26		0.16	0.00	***
Yoon 지영	0.06	0.20		0.03	0.48		0.04	0.43		0.11	0.03	*
Yoon 지희	0.06	0.20		0.03	0.48		0.04	0.43		0.11	0.03	*
Yoon 형오	0.06	0.23		0.03	0.52		0.04	0.44		0.10	0.04	*
Yoon 흥수	0.05	0.29		0.03	0.60		0.03	0.52		0.10	0.05	*
김남윤	0.06	0.23		0.03	0.49		0.04	0.43		0.11	0.03	*
이현지	0.06	0.25		0.03	0.54		0.04	0.48		0.11	0.03	*
윤태민	0.07	0.19		0.04	0.46		0.04	0.42		0.13	0.01	*
한둥근체 돋움	0.13	0.01	**	0.11	0.03	*	0.11	0.03	*	0.21	0.00	***
한둥근체 바탕	0.08	0.09	.	0.06	0.26		0.06	0.21		0.13	0.01	**
KCC 은영	0.28	0.00	***	0.24	0.00	***	0.22	0.00	***	0.39	0.00	***
KCC 김훈	0.18	0.00	***	0.15	0.00	**	0.14	0.01	**	0.18	0.00	***
한글누리	0.08	0.09	.	0.04	0.38		0.04	0.37		0.15	0.00	**
기랑해랑체	0.18	0.00	***	0.12	0.02	*	0.12	0.02	*	0.22	0.00	***

Note. * $p < .05$, ** $p < .01$, *** $p < .001$.

We further investigated the level of contribution of each letter to the whole systematicity by excluding each letter in turn and re-conducting the correlation test. Table 5 shows that when the overall correlation was .3, removing individual letters increased or decreased the correlation accordingly. The consonants individually tend to contribute positively to the whole grapho-phonemic systematicity, whereas the vowels tend to hinder it. For example, without ㄸ and ㅃ, the correlation decreased to $r = .27$ ($p < .001$) whereas excluding ㅡ increased the coefficient to $r = .4$ ($p < .001$).

Table 5. The contribution of each letter in 궁서. Subtracting individual letters changes the overall correlation of .3

excluded item	*r*
ㅃ	0.27
ㄸ	0.27
ㅍ	0.28
ㅌ	0.28
ㅎ	0.28
ㅂ	0.28
ㅉ	0.28
ㅈ	0.28
ㄷ	0.28
ㅁ	0.28
ㅋ	0.28
ㅊ	0.29
ㄲ	0.29
ㅆ	0.29
ㄹ	0.29
ㅇ	0.30
ㅅ	0.30
ㅐ	0.30
ㅔ	0.31
ㅓ	0.31
ㄴ	0.32
ㄱ	0.32
ㅢ	0.33
ㅣ	0.33
ㅏ	0.33
ㅟ	0.34
ㅜ	0.34
ㅗ	0.36
ㅡ	0.40

4. Discussion

Artificially designed with an explicit pedagogical aim, hangeul has a widely known intrinsic systematicity between letter shapes and their pronunciations. We successfully quantified its systematic relation between Korean letter shapes and their sounds and defined it as grapho-phonemic systematicity. Predictably, stroke share rate returned the highest correlation values; strokes reflect higher-level, consciously appreciated structure. However, Hausdorff distance is more cross-linguistically applicable; although it returned slightly reduced correlation values, it was still robust and has the advantage of being able to reveal unappreciated contributions to systematicity. Hangeul, the result of deliberate cultural invention, is the gold standard of grapho-phoneme systematicity among scripts.

With Hausdorff distance, we have demonstrated grapho-phonemic systematicity in Arabic, Cyrillic, English, Finnish, Greek, and Hebrew (Jee, Tamariz, & Shillcock, 2021; 2022a; 2022b). These scripts returned in general positive grapho-phonemic systematicity, indicating that similar letters tend to have similar sounds, or vice versa. In the most recent finding, Chinese characters also showed a positive syllable-character systematicity (Du et al., 2022; Jee et al., 2022b). However, none of them showed higher systematicity than hangeul.

The consistent variation of the letter forms seen in hangeul is also found in Hebrew, which also, shows comparatively robust grapho-phonemic systematicity across fonts (Jee et al., 2022a). As in hangeul, Hebrew varies the letter shapes according to the voiced-voiceless contrast (e.g., ב /v/ - בּ /b/) and the manner of articulation (e.g., voiceless velar plosive ךּ /k/ - voiceless velar fricative ך /x/).

There are two ways to vary letter shape consistently at the level of the whole alphabetic system: (i) add or subtract letter elements; or (ii) change orientation of the identical letter. In conventional orthographies (Hebrew, Burmese, Runic, and even cuneiform) the former is preferred. In some artificial scripts (e.g., the Shavian alphabet) the latter also occurs. There seem to be several reasons for this, all concerned with kinetic efficiency and least effort (Zipf, 1949/2016).

First, adding a stroke or dot may be kinetically easier than changing orientation. Efficiency is particularly important for high frequency letters. Just as frequent phonemes have reduced distinctiveness (Gahl et al., 2012; Meylan & Griffiths, 2017; Shi et al., 1998), we assume that high frequency letters have simpler letter shapes. This selective pressure was realized as diacritics, for example, in Arabic and Hebrew vowels and even omission of vowels in the unpointed script.

Second, changing orientation affects the general direction of letter faces of an alphabet set (Watt, 1994). Facing direction is defined as the direction of ornaments and headings: for example, Arabic numbers mostly face left. Watt (1994) claims that we are sensitive to the particular asymmetry of the set of letters; children often reverse “b” until they understand that asymmetric Roman letters generally face rightwards.

Finally, letter shapes have implications for writing position and writing time. With pens and pencils, there exist opposite pressures between the three writing fingers and the two supporting fingers. Depending on writing direction, these two pressures alternate the active and passive roles (Watt, 1994). Therefore, it is plausible that cursive scripts optimize the balance between these kinetic forces for the sake of writing speed. Orientational change of letters may hinder writing speed by varying the starting points of letters.

Some fonts returned higher correlation coefficients than others, which indicate that they may emphasize the phonemic regularity. This implies there are pedagogical implications for beginning readers. We are currently investigating the behavioural consequences of grapho-phonemic systematicity at the letter level.

5. Conclusion

We have developed what we take to be the first method for quantifying in a detailed way the systematicity between letter shapes and their corresponding sounds. The method is general enough to be applied to any phonographic orthography. It compares well with a hangeul-specific method based on shared strokes. The method allows us to begin studying the behavioural consequences of grapho-phonemic systematicity, with hangeul emerging as the writing system in which this systematicity is clearest—the gold standard of systematicity.

References

Briggs, R., & Hocevar, D. J. (1975). A new distinctive feature theory for upper case letters. *Journal of General Psychology, 93*(87), 87-93. https://doi.org/10.1001/archopht.1975.01010020091017

Dautriche, I., Mahowald, K., Gibson, E., & Piantadosi, S. T. (2017). Wordform similarity increases with semantic similarity: An analysis of 100 languages. *Cognitive Science*, *41*(8), 2149-2169. https://doi.org/10.1111/cogs.12453

Du, F., Jee, H., Tamariz, M., & Shillcock, R. (2022). Grapho-Syllabic Systematicity in Chinese: Chinese Pictographs Have a Non-Arbitrary relation with their Pronunciations. In *Proceedings of the Annual Meeting of the Cognitive Science Society* (Vol. 44, No. 44). Retrieved from https://escholarship.org/uc/item/5817534p

Gahl, S., Yao, Y., & Johnson, K. (2012). Why reduce? Phonological neighborhood density and phonetic reduction in spontaneous speech. *Journal of Memory and Language*, *66*(4), 789-806. https://doi.org/10.1016/j.jml.2011.11.006

Geyer, L. H. (1977). Recognition and confusion of the lowercase alphabet. *Perception & Psychophysics*, *22*(5), 487-490. https://doi.org/10.3758/BF03199515

Haralambous, Y. (2021, February). *Proceedings of Grapholinguistics in the 21st Century*, June 17-19, 2020. In G21C 2020: *Grapholinguistics in the 21st Century* (Vol. 4). Fluxus Editions.

Hideki, N. (2011). *한글의 탄생; 문자라는 기적*. Gyeongido Paju: Dolbaegae.

Huttenlocher, D. P., Klanderman, G. A., & Rucklidge, W. J. (1993). Comparing images using the Hausdorff distance. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *15*(9), 850-863. https://doi.org/10.1109/34.232073

Hyun, P. (1981). A trove of unfamiliar art from Korea. *New York Times*, *4*, d1.

Jee, T., & Shillcock. (2020). Quantifying sound-graphic systematicity and application on multiple phonographs. *Cognitive Science Proceedings.*

Jee, H., Tamariz, M., & Shillcock, R. (2021). Quantifying sound-graphic systematicity; Application to multiple phonographic orthographies. Haralambous, Y. (2021, February). *Proceedings of Grapholinguistics in the 21st Century*, June 17-19, 2020. In G21C 2020: *Grapholinguistics in the 21st Century* (Vol. 4). Fluxus Editions.

Jee, H., Tamariz, M., & Shillcock, R. (2022a). Systematicity in language and the fast and slow creation of writing systems: Understanding two types of non-arbitrary relations between orthographic characters and their canonical pronunciation. *Cognition*, *226*, 105197. https://doi.org/10.1016/j.cognition.2022.105197

Jee, T., & Shillcock. (2022b). Letters and their sounds are not perfectly arbitrary: Exploring grapho-phonemic systematicity in multiple orthography systems. *Proceedings of the Joint Conference on Language Evolution (JCoLE) Evolution of Language*, pp. 354-356.

Martin, A., Kronbichler, M., & Richlan, F. (2016). Dyslexic brain activation abnormalities in deep and shallow orthographies: A meta-analysis of 28 functional neuroimaging studies. *Human Brain Mapping*, *37*(7), 2676-2699. https://doi.org/10.1002/hbm.23202

Meylan, S. C., & Griffiths, T. L. (2017). Word forms-not just their lengths-are optimized for efficient communication. *arXiv preprint arXiv:1703.01694*.

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. https://doi.org/10.1098/rstb.2013.0299

Paulesu, E., Démonet, J. F., Fazio, F., McCrory, E., Chanoine, V., Brunswick, N., ... & Frith, U. (2001). Dyslexia: cultural diversity and biological unity. *Science*, *291*(5511), 2165-2167. https://doi.org/10.1126/science.1057179

Paulesu, E., McCrory, E., Fazio, F., Menoncello, L., Brunswick, N., Cappa, S. F., ... & Pesenti, S. (2000). A cultural effect on brain function. *Nature neuroscience*, *3*(1), 91-96. https://doi.org/10.1038/71163

Sampson, G. (1985). *Writing systems.* London, UK: Hutchinson.

Seymour, P. H., Aro, M., Erskine, J. M., & Collaboration with COST Action A8 Network. (2003). Foundation literacy acquisition in European orthographies. *British Journal of psychology*, *94*(2), 143-174. https://doi.org/10.1348/000712603321661859

Shi, R., Morgan, J. L., & Allopenna, P. (1998). Phonological and acoustic bases for earliest grammatical category assignment: A cross-linguistic perspective. *Journal of Child Language*, *25*(1), 169-201. https://doi.org/10.1017/S0305000997003395

Spencer, K. (2001). Differential effects of orthographic transparency on dyslexia: Word reading difficulty for common English words. *Dyslexia*, *7*(4), 217-228. https://doi.org/10.1002/dys.207

Tamariz, M. (2008). Exploring systematicity between phonological and context-cooccurrence representations of the mental lexicon. *The Mental Lexicon*, *3*(2), 259-278. https://doi.org/10.1075/ml.3.2.05tam

Watt, W. C. (1979). Iconic equilibrium. *Semiotica, 28*(1-2), 31-62. https://doi.org/10.1515/semi.1979.28.1-2.31

Watt, W. C. (1994). *Writing Systems and Cognition. Perspectives from Psychology, Physiology, Linguistics, and Semiotics*. Springer: Netherlands. https://doi.org/10.1007/978-94-015-8285-8

Zipf, G. K. (1949/2016). *Human behavior and the principle of least effort: An introduction to human ecology*. Ravenio Books.

Note

Note 1. The Python code and data are available from https://github.com/HanaJee/hausdorff-distance-letters.git.

Copyrights