

Sound-graphic Systematicity Quantified and Found in Multiple Phonographs

Dr. Hana Jee, York St John University

Dr. Monica Tamariz, Heriot-Watt University

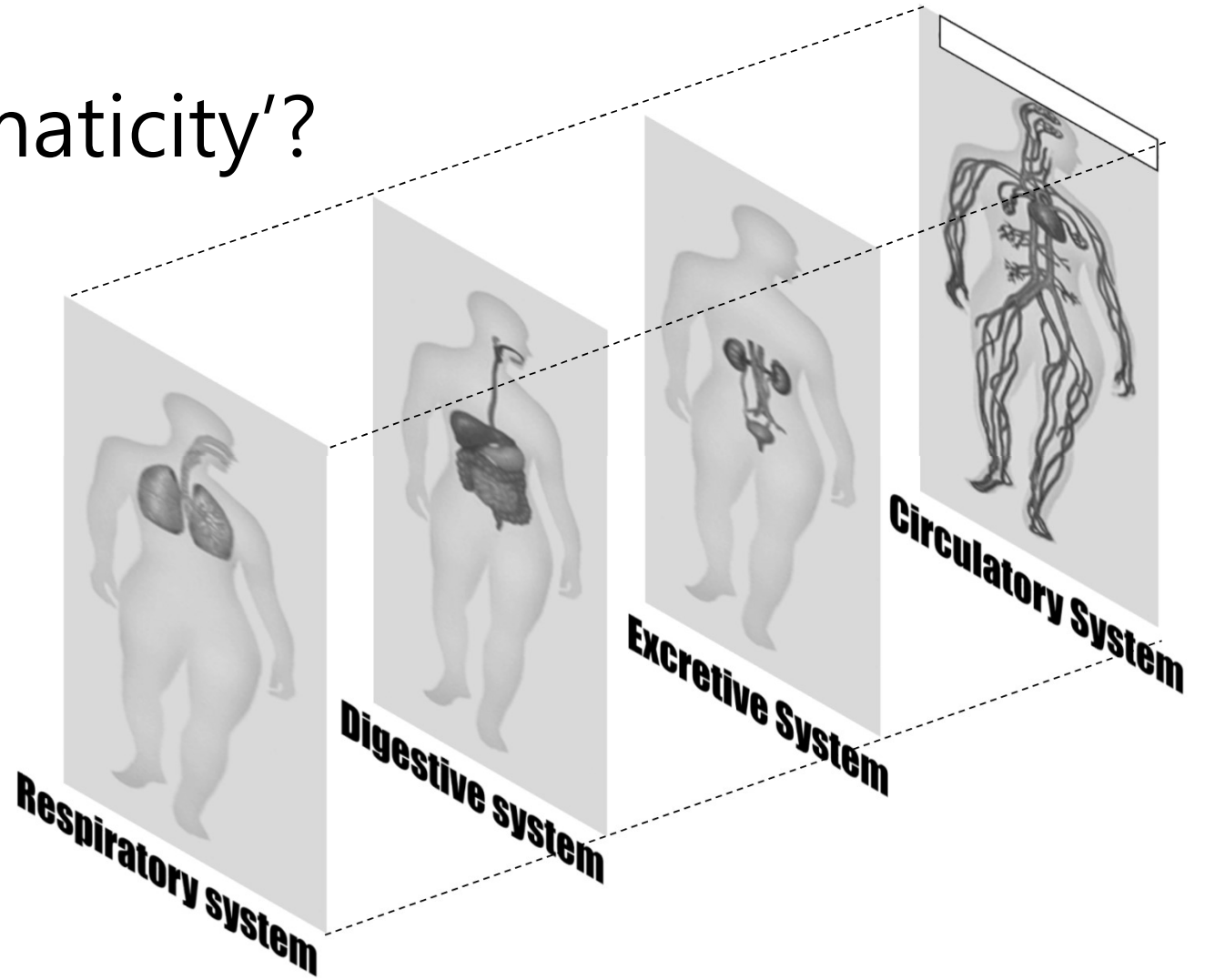
Dr. Richard Shillcock, The University of Edinburgh

Jee, H. (2021). *Linguistic systematicity between phonology, semantics, and orthography in English and Korean*. (Doctoral dissertation, The University of Edinburgh, Edinburgh, UK).

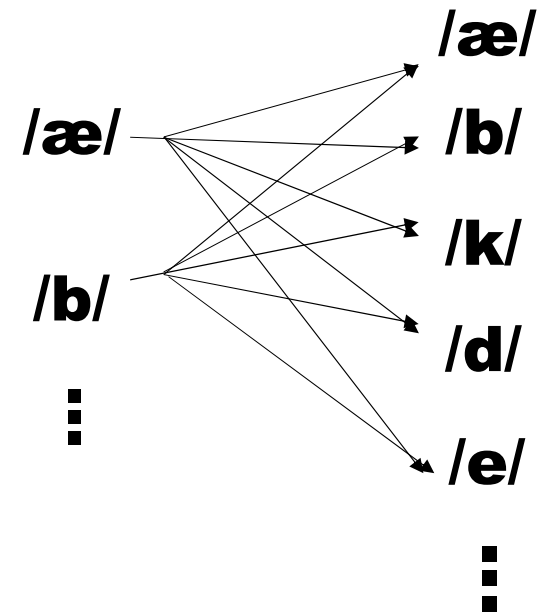
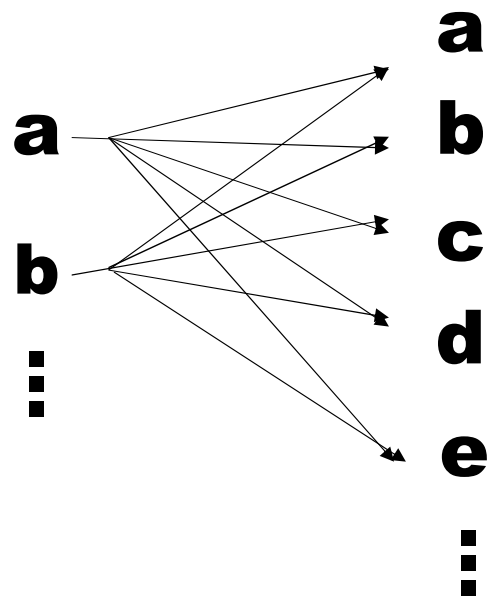
Language as a Complexity System

- Large number of agents
- Interaction among the agents
- Nonlinear, disproportionate effect
- Unpredictable
- Dynamic systems
- Sensitive to the initial condition butterfly effect

What is 'Systematicity'?



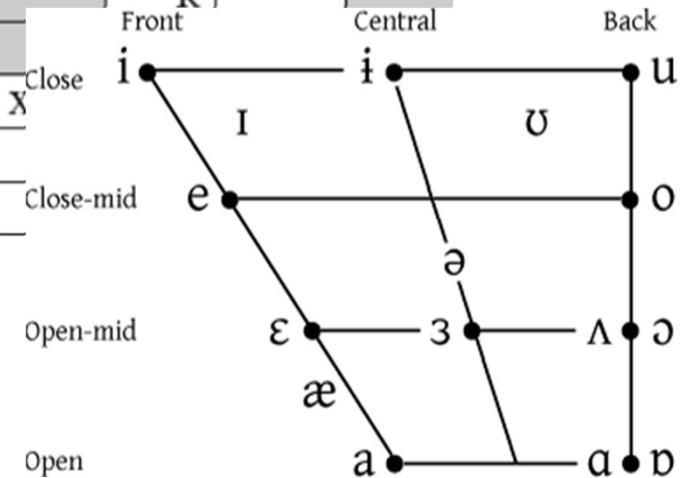
"LETTERS are
NOT perfectly ARBITRARY
with SOUNDS."



Tamariz 2008; Monaghan, Shillcock, Christiansen, & Kirby, 2014

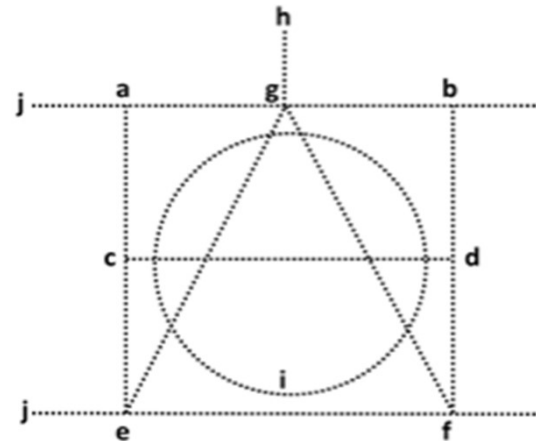
How to Measure Phonological Distances

	Bilabial	Labiodental	Dental	Alveolar	Postalveolar	Retroflex	Palatal	Velar	Uvular	Pharyngeal	Glottal
Plosive	p b			t d		ʈ ɖ	c ɟ	k ɡ	q ɢ		ʔ
Nasal	m	ɱ		n		ɳ	ɲ	ŋ	ɴ		
Trill	ʙ			ʀ					ʀ		
Tap or Flap		ⱱ		ɾ		ɽ					
Fricative	ɸ β	f v	θ ð	s z	ʃ ʒ	ʂ ʐ	ç ʝ	x			
Lateral fricative				ɬ ɮ							
Approximant		ʋ		ɹ		ɻ	j				
Lateral approximant				l		ɭ	ʎ				



How to Measure Orthographical Distances

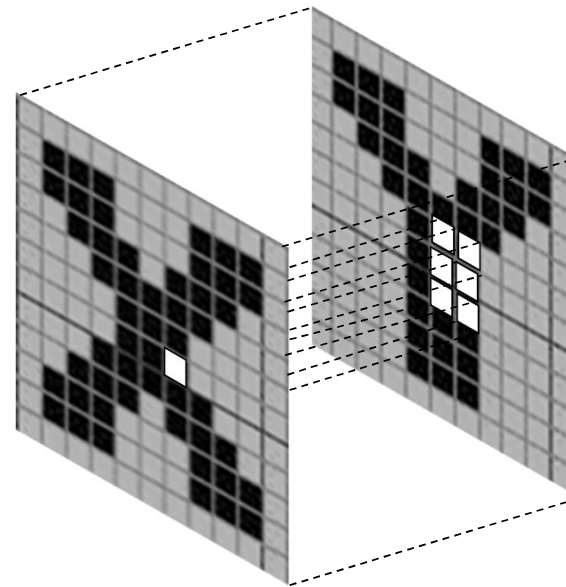
1. Pixel count
2. Perimetric Complexity
Pelli, et al., 2006
3. Stroke Share Rate
4. Hausdorff Distance
Huttenlocher, et al., 1993



	a-c	d-e	g-h	c-e		dup
⌈	1	0	0	1		0
⌋	1	1	0	1	0
⌈⌋	1	0	0	1		1

How to Measure Orthographical Distances

1. Pixel count
2. Perimetric Complexity
Pelli, et al., 2006
3. Stroke Share Rate
4. Hausdorff Distance
Huttenlocher, et al., 1993



Letter 1	Letter 2	Font	Hausdorff <i>Max</i>
a	b	Courier New	2.65
		Georgia	2.45
		Times New Roman	2.24

بغرافية فوالحي
 متى الاستقرار فنث
 معاومة اعزاء مختلف
 المغربية بما كان تيسر فيها
 اذ ظهرت آثارها في بعض أعمال
 ستان المعماري الليبي استطاعت أن
 حمال ولباطة المشيد
 الله باشا» بأستنبول
 ليبيا العديدين
 الليبي

Надкъ примѡрога чкнрч г

Съ погомъ повнѣмо слопо примѡрога
 слопн. снѣгнрнч царь шнеше оАорсн
 незнкъ. чкнръ примѡрн кнжннкъ с
 номъ спономъ ъ лнепч, н те нмаше Анеп
 нч спакн вчсь сч пларгоѡфннеъмъ н с пел
 ѡкчзѡне спото мѡкѡ н кѡнѡ, н реве пемѡ
 прошеннѣ нспѡннхъ, ч цю не ѡ Анеце п
 мепе; ѡ пепе несць сестрнннѣ чнччннѣ.
 кчкко снѣч спогч.» н реве чкнръ: «ѣте спож
 те спно мѡшкн плоАь, ѡснччо спн парспть
 смѡрпн монѣ. чкко лн спн кнн спнлч, пѡсп

inen johon luotat ihan helveti
 on ja luulet että se ei satutta
 koskaan, satuttaa sua koko
 vaan enemmän ja enemmän
 ää sun luottamuksen..

hey travelled by plane and stayed in
 g holiday, and the first problem was
 didn't take skis, sticks, boots or goggles
 Moritz. She took a pair of sunglasses,
 even pairs of warm socks with her. T
 me. The sky was blue and clear, but it
 skiing and she fell over hundreds of
 r ski instructor, but people often fall

ἀπελευθέρωσιν καὶ εἰς τοὺς τυφλοὺς
 ἐκείνους ποὺ πιέζονται, νὰ κηρύξῃ
 Κυρίου. Καὶ ἀφοῦ ἐτύλιξε τὸ βιβλ
 ὑπηρέτην, ἐκάθησε καὶ τὰ μάτια ὄλ
 προσηλωμένα ἐπάνω του. Ἄρχισε
 ἔχει ἐκπληρωθῆ εἰς τὰ αὐτιά σας ἡ
 κατένευαν πρὸς αὐτὸν καὶ ἐθαύμαζ
 ποὺ ἔβγαϊναν ἀπὸ τὸ στόμα του ἡ
 ὁ υἱὸς τοῦ Ἰωσήφ;”. Καὶ ὁ Ἰησο
 οῦ πῆτε αὐτὴν τὴν παροιμίαν,
 ου. Ὅσα ἀκούσαμε ὅτι ἔγιναν εἰ
 αὶ ἐδῶ εἰς τὴν πατρίδα σου”. Κα
 γω ὅτι κανένας προφήτης δέν

4 קדשו ישמח לבו מבקשי יהוה: דרשו יהוה
 ה פניו תמיד: זכרו נפלאותיו אשר-עשה מפ
 6 פיו: זרע אברהם עבדו בני יעקב בחירו
 8 אלהינו בכל-הארץ משפטיו: זכר לעולם
 9 צנה לאלה דור: אשר-ברת את-אברהם ושבע
 11 ונעמדה ליעקב לחק לישראל ברית עולם:
 12 אתן את-ארץ כנען חבל גחלתכם: בהיותם מ
 13 כמעט ונרים בה: ויתהלכו מגוי אל-גוי מממלכו
 14 אחר: לאהגית אדם לעשקם ויובח עליהם
 16 אל-תגעו במשיחי ולגביאי אל-תרעו: ויקרא ר
 17 הארץ כל-מטה-לחם שבר: שלח לפניהם איש
 18 נמבר יוסף: ענו בכבל רגליו ברזל באה נפשו: איש
 19 באדברו אמרת יהוה צרפתחו: שלח מל
 משל עמים ויפתחו: שמו א



































Ancient Semitic Orthographies

⋆	'	⊗	T	⤴	P
⤴	B	⤵	Y	⤵	S
⤵	G	⤴	K	⤴	Q
⤴	D	⤵	L	⤴	R
⤴	H	⤴	M	⤴	Š
⤴	W	⤴	N	⤴	T
⤴	Z	⤴	S		
⤴	H	○	'		




1	⋆	⤴	⤵	⤴	⤴	⤴	⤴	⊗	⤴	⤵	⤵	⤴	⤴	⤴	⊗	○	⤴	⤴	⤴	⤴	⤴	⤴
2	⤵	⤴	⤴	⤴	⤴	⤴	⤴	⤴	⤴	⤴	⤴	⤴	⤴	⤴	⤴	⤴	⤴	⤴	⤴	⤴	⤴	⤴
3	ا	ب	ج	د	هـ	و	ز	ح	ط	ي	ك	ل	م	ن		ع	ف	ص	و	ر	س	ش
4	⤵	⤴	⤴	⤴	⤴	⤴	⤴	⤴	⤴	⤴	⤴	⤴	⤴	⤴	⤴	⤴	⤴	⤴	⤴	⤴	⤴	⤴
5	'	b	g	d	h	w	z	ḥ	ṭ	y	k	l	m	n	s	'	p/f	ṣ	q	r	š	t
1. Aramaic ; 2. Nabataean ; 3. Arabic ; 4. Syriac ; 5. Transcription																						

Fictitious Orthographies



					
Aurek (a)	Besh (b)	Cresh (c)	Cherek (ch)	Dorn (d)	Esk (e)
					
Enth (ae)	Onith (eo)	Forn (f)	Grek (g)	Herf (h)	Isk (i)
					
Jenth (j)	Krill (k)	Krenth (kh)	Leth (l)	Mern (m)	Nern (n)
					
Nen (ng)	Osk (o)	Orenth (oo)	Peth (p)	Qek (q)	Resh (r)
					
Senth (s)	Shen (sh)	Trill (t)	Thesh (th)	Usk (u)	Vev (v)
					
Wesk (w)	Xesh (x)	Yirt (y)	Zerek (z)		



									
a	b	ch	D	e	gh	H	I	l	j
									
l	m	n	ng	o	p	q	Q	r	
									
S	t	th	u	v	w	y	y	y	y
									
0	1	2	3	4	5	6	7	8	9

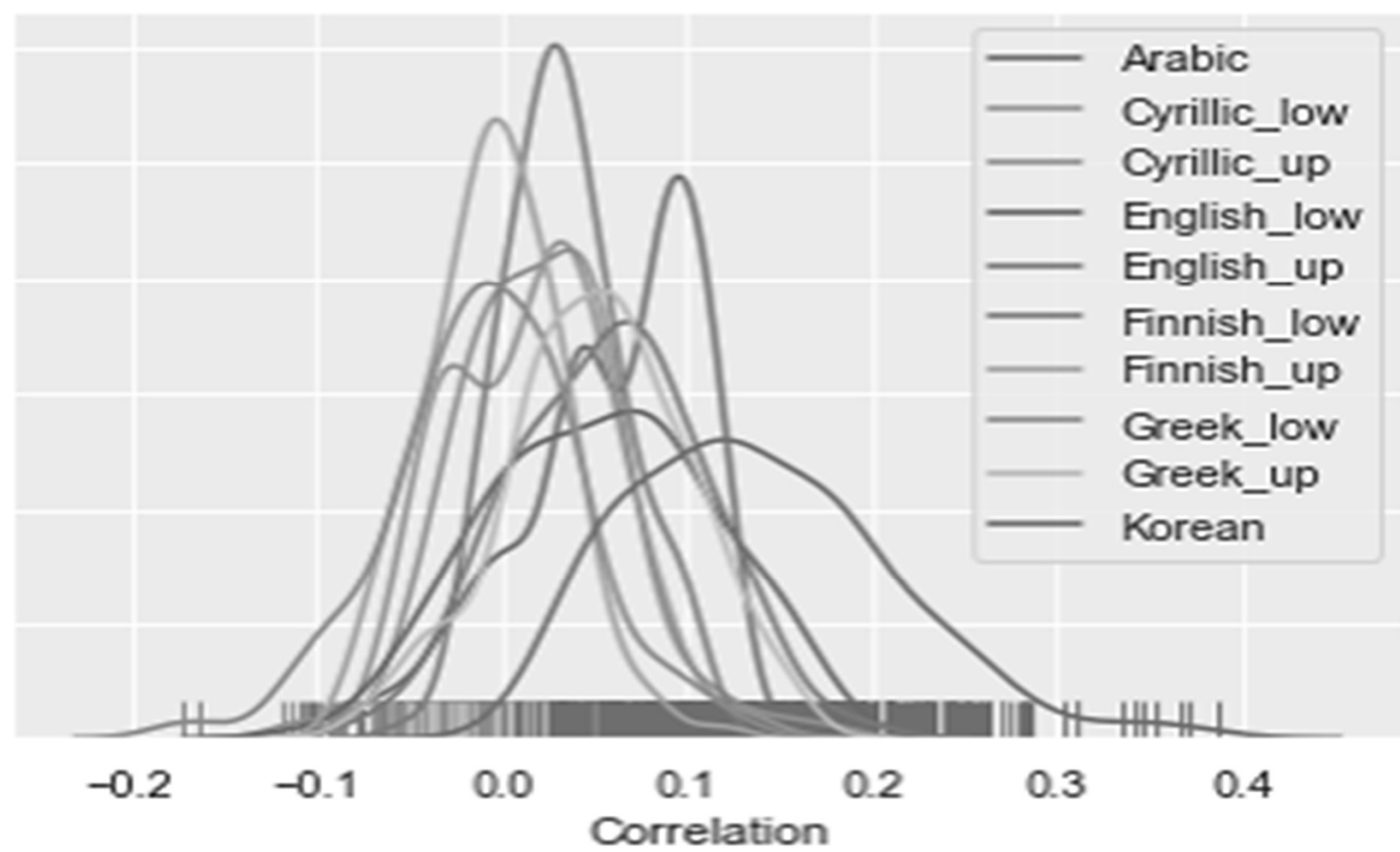
1. Non-Arbitrary Letter-Sound Relation

Non-Arbitrary

- Natural Orthographies
 - Arabic, Cyrillic, English, Finnish, Greek, Hebrew, and Chinese
- Artificial Orthographies
 - Korean, Pitman's shorthand, Shavian alphabet

Arbitrary

- Fictitious Orthographies
 - Ancient Orthographies
-



2. Two Types of Systematicity

Effort-Based

- Pixel Count
 - ink area
- Perimetric Complexity
 - the number of strokes

Arabic, Cyrillic, English, Greek, Hebrew

- Higher frequency induces simplicity
- Symmetric motoric effort between hands and mouth

Symbolic

- Hausdorff Distance
 - image similarity

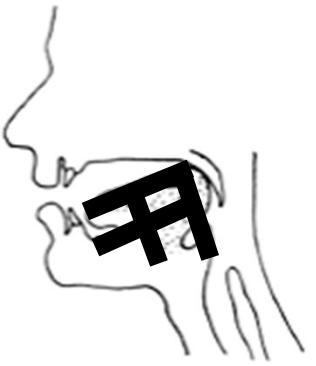
Korean, Pitman's, Shavian

- Similar phonemes designed visually similar
-

Source of Symbolic Systematicity



/g/ /n/ /m/ /s/ /ng/



└	/g/	guy	└	└	└	└
	/g/		/g/	/d/	/b/	/dg/
≡	/k/	<u>key</u>	≡	≡	≡	大
	/k/		/k/	/t/	/p/	/ch/
└└	/k*/	sk <u>y</u>	└└	└└	≡≡	└└
	/k*/		/k*/	/t*/	/p*/	/tch/

Source of Symbolic Systematicity

/p/	
/b/	Ⓛ
/t/	1
/d/	Ⓟ
/tʃ/	z
/dʒ/	2
/k/	Ⓞ
/g/	Ⓟ
/f/	J
/v/	f
/θ/	
/ð/	

Shavian alphabet

	<i>r</i>	<i>p</i>	
Euclid	0.23	0.00	***
Edit	0.23	0.00	***

Pitman shorthand

	<i>r</i>	<i>p</i>	
Euclid	0.12	0.05	*
Edit	0.14	0.02	*

/p/	
/b/	Ⓛ
/t/	
/d/	
/tʃ/	/
/dʒ/	/
/k/	—
/g/	—
/f/	⌒
/v/	⌒
/θ/	
/ð/	

2. Two Types of Systematicity

Effort-Based

- Pixel Count
 - ink area
- Perimetric Complexity
 - the number of strokes

Arabic, Cyrillic, English, Greek, Hebrew

- Higher frequency induces simplicity
- Symmetric motoric effort between hands and mouth

Symbolic

- Hausdorff Distance
 - image similarity

Korean, Pitman's, Shavian

- Similar phonemes designed visually similar
-

3. Method designed for the language

Hausdorff Distance

- Purely computational
- Applicable to many language

$$r = .39$$
$$p < .001$$

Stroke Share Rate

- Designed for Korean
- Maximize the visual systematicity

$$r = .66$$
$$p < .001$$

4. Some Fonts have stronger systematicity

	Font	<i>r</i>	<i>p</i>	
1	KCC 은영	0.39	< .0001	***
2	HS가을생각체	0.37	< .0001	***
3	KBIZ한마음고딕	0.37	< .0001	***
4	HS두꺼비체	0.35	< .0001	***
5	KBIZ한마음명조	0.35	< .0001	***
..... ellipsis				
83	Yoon흥수	0.1	0.05	*
84	한겨레결체	0.08	0.13	
85	조선일보명조체	0.08	0.09	.
86	FB이철수2001목판M	0.08	0.11	
87	FB이철수2001목판TM	0.08	0.11	
88	신비는일곱살	0.06	0.26	

	r	p	
Georgia	0.273	0.000	***
Book Antiqua	0.262	0.000	***
Palatino Linotype	0.260	0.000	***
Constantia	0.258	0.000	***
Candara	0.250	0.000	***
Cambria	0.249	0.000	***
Times New Roman	0.231	0.000	***
Trebuchet MS	0.227	0.000	***
Calibri Light	0.209	0.000	***
Calibri	0.207	0.001	***
Corbel	0.203	0.001	***
Franklin GothicMedium	0.200	0.001	***
<i>Segoe Print</i>	0.200	0.001	***
Lucida Sans Unicode	0.199	0.001	***
Tahoma	0.198	0.001	***
Century Gothic	0.187	0.002	**
Comic Sans MS	0.182	0.002	**
Microsoft Sans Serif	0.181	0.002	**
Arial	0.175	0.003	**
Verdana	0.173	0.004	**
Segoe UI Symbol	0.172	0.004	**
Arial Black	0.162	0.007	**
Consolas	0.151	0.012	*
<i>Segoe Script</i>	0.151	0.012	*
Gabriola	0.147	0.015	*
Courier New	0.136	0.023	*
Impact	0.136	0.024	*
Lucida Console	0.098	0.103	
<i>Lucida Handwriting</i>	0.084	0.164	

Summary

1. Letters are not perfectly arbitrary with their sounds.
2. Two types of the systematicity: effort-led vs. symbolic
3. Method designed for the language maximizes the systematicity.
4. Some fonts inhere higher letter-sound systematicity.

Conclusion

- Brain maximizes systematicity.
 - Language = full of arbitrariness
 - Principle of least effort (Zipf, 1949)
 - Statistical Learning

- Letter-Sound connection matters.
 - Literacy Rate caused by orthographic depth
 - Shallow orthography vs. Deep orthography

- High correlation can help...
 - Children who first learn sound-letter relation
 - Deep orthography users
 - The dyslexic with certain type



Thank you!