Phonology of gender in Korean personal names

Yoonjung Kang & Lisa Sullivan
University of Toronto

June 5, 2021

Korean Linguistics in Crosslinguistic Context

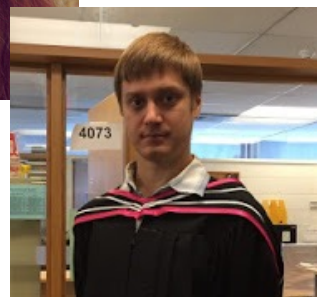
Overview

Questions: Are there cross-linguistic tendencies of non-arbitrary connection between gender and phonological feature in personal names? How do those tendencies interact with language-specific sound symbolism?

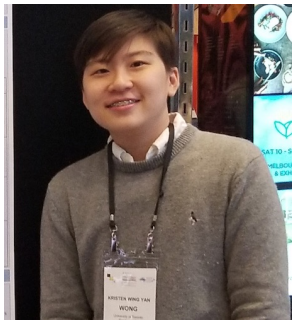
Outline

1. Introduction
2. Cross-linguistic survey
3. Korean dark & light vowels
4. Conclusion

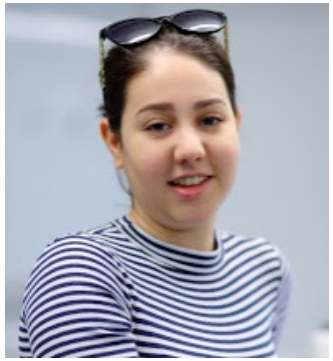
Contributors

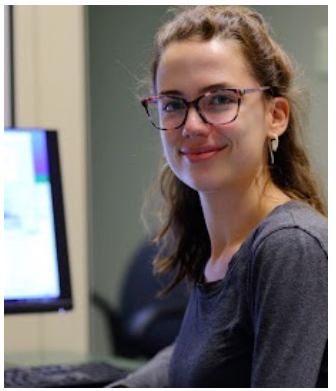


- Lisa Sullivan – **Can. Eng, Can. French**, **Korean**
- Andrei Munteanu – **Russian**
- Kristen Wong – **Cantonese**
- Nazia Mohsin – **Urdu**
- Gajathree Ananthathurai – **Tamil**



- Gajathree Ananthathurai, Laurestine Bradford, Araz Derohan, Siobhan Galeazzi, Khadija Jagani – **Armenian, Kutchi**

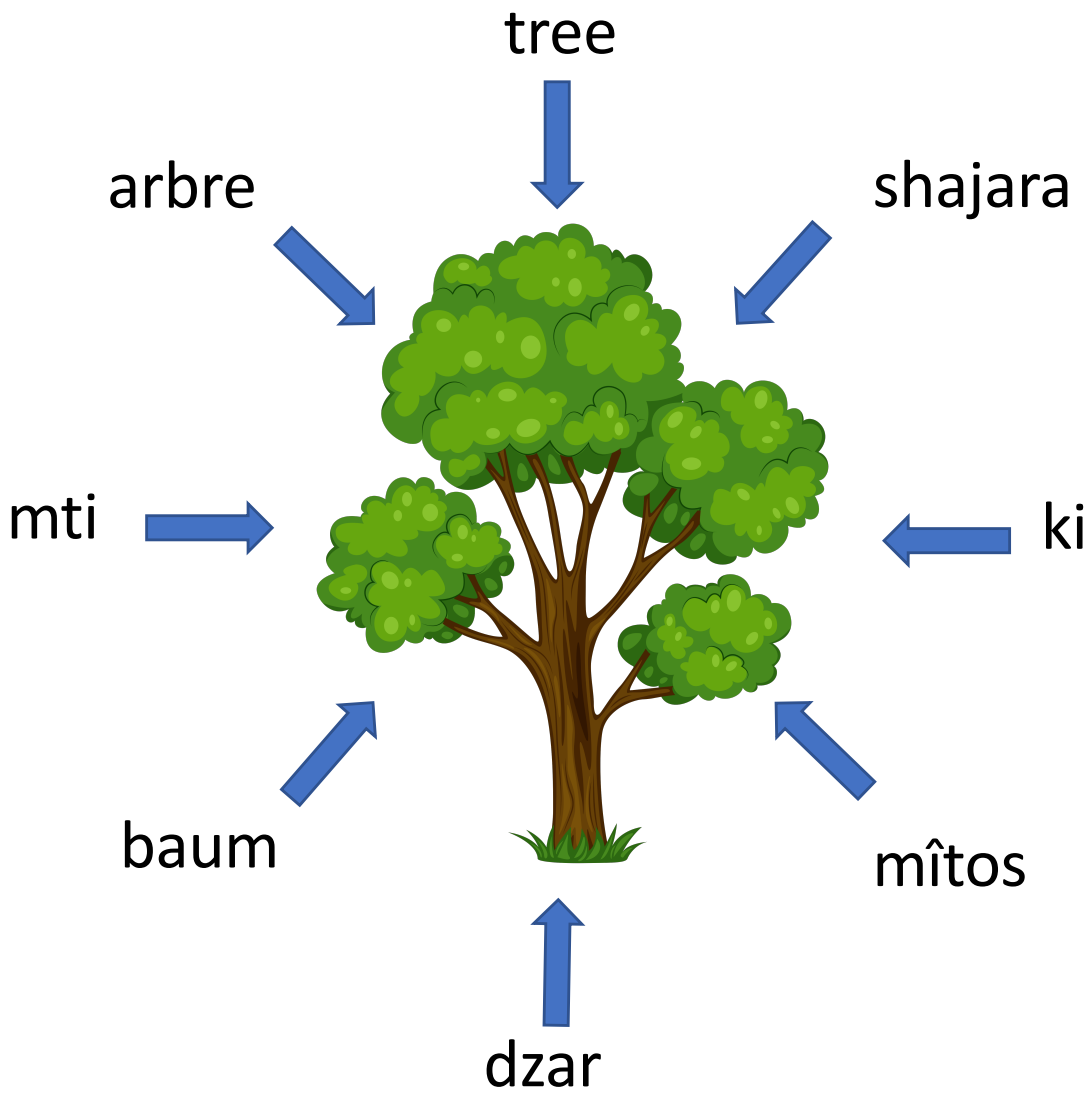




Introduction

The arbitrariness of the sign

- The form or sound of a word does not resemble its meaning (Saussure 1916).
- Considered to be one of the major design features of human language (Hockett 1960).


tree
arbre
shajara
mti
ki
baum
mîtos
dzar

Non-arbitrariness

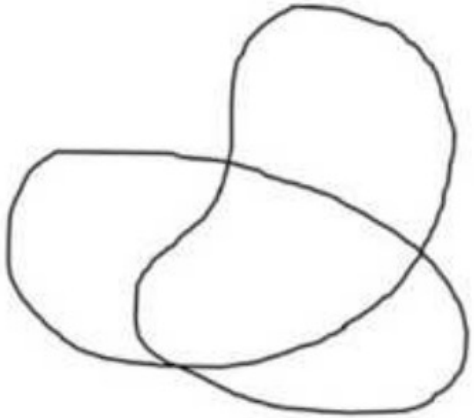
- Non-arbitrary connections between sound and meaning/word class are more pervasive than generally assumed.

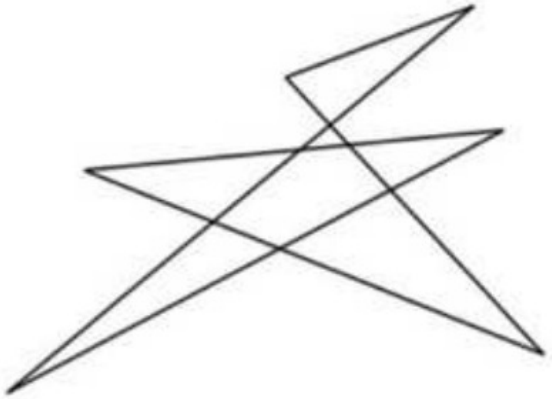
(Kohler 1922; Sapir 1929; Newman 1933;Ultan 1978; Kim-Renaud 1978; Kawahara & Shinohara 2012; Dingemanse et al., 2015 ; Monaghan, 2014; Svantesson 2016; Fowler 2016; Alderete and Kochetov 2017; Shih et al. 2019)

- ***Systematic***: some sound-meaning connections are accidental statistical tendencies particular to specific languages or arise from culture-specific social meanings, with no expectations for cross-linguistic generalizability.
- ***Iconic***: other connections may have perceptuo-motoric grounds that make them more amenable to sound-meaning association.

(Monaghan et al,. 2014)

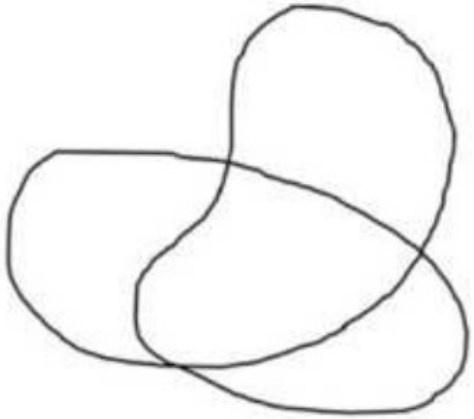
Sound symbolism

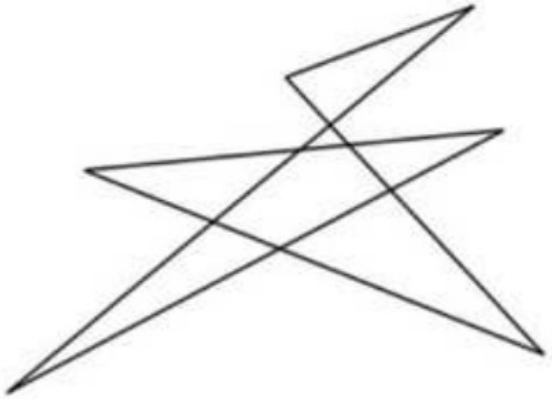




Maluma Takete

Sound symbolism





- Maluma-Takete/Bouba-Kiki Experiment (Köhler 1929)
  - ~95-98% said: round shape is "bouba", spiky shape is "kiki" (Ramachandran & Hubbard, 2001)
  - Same results with adults and 2.5 year old children (Maurer, Pathman, & Mondloch, 2006)

Sound symbolism in names

- Sound symbolism has been explored productively in names.
  - Pokemons, baseball players, product names, aliens, superheroes, personal names

English personal names

- Many studies on English names.
- Correlation between sound patterns and gender observed.

  Mary vs. John

- Female names:
  - tend to be longer (in terms of syllable number)
  - sonorants more common in beginnings and endings
  - tend to end in an open syllable
  - prefer front vowels
- Male names:
  - More obstruents and stop endings
  - Tend to end in a closed syllable
  - Prefer back vowels

Cassidy, Kelly & Sharoni, 1999; Cutler, McQueen & Robinson, 1990; Pitcher, Mesoudi & McElligott, 2013; Sidhu & Pexman, 2015; Slater & Feinman, 1985; Sutton, 2016; Sullivan 2018; Shih 2019.

Beyond English

- Some of the patterns may have plausible sound symbolic grounds.
- The sound symbolism may not directly signal gender, but relates to concepts stereotypically linked to gender, e.g., size.
- Do other languages exhibit gender-phonology connection in names? How? Are the patterns found in English also attested in other languages?
- Studies outside English names are rare.
  - But, see Starr, et al. 2018 on Cantonese and Mandarin and Shinohara & Kawahara 2013 on Japanese.

Crosslinguistic studies

Languages and data sources

- Languages were selected opportunistically
- Data sources
  - Government registry/survey data
    - Russian (356), Korean (1938), Canadian French (199), Cantonese (288)
  - School yearbook, student list, membership list, names dictionary
    - Armenian (198), Tamil (400), Kutchi (386)
  - Names elicited from native speakers
    - Urdu (200)

Predictions: Vowels

- “The Frequency Code” (Ohala 1983, 1984, 1994)
  - Sounds with high acoustic frequency (front unrounded vowels, high tone, and acute consonants) are used in words denoting smallness

    (Kawahara & Shinohara, 2013; Newman, S. S. 1933 ; Ohala, J. J. 1983; 1984; 1994; Pitcher, Mesoudi & McElligott, 2013)
  - Front vowels, specially high front /i/, often associated with femininity in sound symbolism

    (Cutler, McQueen & Robinson, 1990; Miall, 2001; Wu, Klink & Guo, 2013)
- Articulatory configuration
  - Wider aperture of low vowels is associated with larger size.

    (Kawahara & Shinohara 2010, Sapir 1929, Jesperson 1922)

Female ~ Front/unrounded, high vowels

Male ~ Back/rounded, low vowels

Predictions: Consonants

- Acoustic transition
  - Obstruents, with abrupt acoustic transition, are associated with angular shape and masculinity while sonorants, with gradual transition, are associated with round shape and femininity. (Shinohara & Kawahara, 2013; Newman, 1933; Sidhu & Pexman, 2015; Shinohara & Kawahara, 2013)
  - But, sonorants also signal weight in Pokemons and baseball player names (Shih and Rudin 2019, Shih et al. 2020)
- "The Frequency Code" (Ohala 1994)
  - Acute consonants (coronals, palatals) with higher frequency noise and formant transitions signal smallness compared to grave consonants (labials, back velars) with lower frequency noise and formant transitions.

Female ~ Sonorants, acute consonants

Male ~ Obstruents, grave consonants

Predictions: Syllable weight

- English names
  - Female names have a higher ratio of open-to-total syllables than male names

  (Slater & Feinman, 1985; Wright, Hay & Bent, 2005)
  - Closed syllables and heavy syllables will favour male names

Female ~ Open/light syllable

Male ~ Closed/heavy syllable

Predictions: word length

- English names
  - Female names are longer in syllables (Cutler, et al. 1990, Slater & Feinman 1985).
- Japanese names
  - *Pokemon* names, length is positively correlated with size, weight, and power (Shih et al. 2020).
  - Personal names, male names are longer in mora count (Mutsukawa 2016).

Female ~ ?

Male ~ ?

Statistics

- Univariate logistic regression for each predictor
  - Dependent variable : gender (F, M)
  - Predictors: phonological features
  - Random effects of syllable/character added for Korean & Cantonese

Results

(with/without gender suffix [-a])

	ENG	FRN	ARM	CAN	KOR		RUS	TAM	KUT	URD
High V	F	ns	F	F	ns		F	F	M	ns
Low V	M	ns	M	M	ns		F/ns	F/ns	F	ns/M
Front V	F	F	F	ns	F		F	F	ns/F	ns
Back V	M	M	M	ns	M		M	M	ns	ns
Son. C	F	F	F	F	F		F	ns/F	ns	ns
Heavy Syl	M	M	NA	M	M		M/ns	NA	NA	M
Longer name	F	F	F	NA	NA		F/ns	F	ns/M	F/M

Results

(with/without gender suffix /-a/)

	ENG	FRN	ARM	CAN	KOR		RUS	TAM	KUT	URD
High V	F	ns	F	F	ns		F	F	M	ns
Low V	M	ns	M	M	ns		F/ns	F/ns	F	ns/M
Front V	F	F	F	ns	F		F	F	ns/F	ns
Back V	M	M	M	ns	M		M	M	ns	ns
Son. C	F	F	F	F	F		F	ns/F	ns	ns
Heavy Syl	M	M	NA	M	M		M/ns	NA	NA	M
Longer Word	F	F	F	NA	NA		F/ns	F	ns/M	F/M

- Vowel height effects are generally consistent with expected tendencies – high vowels prefer Female more than low vowels, except in Kutchi.
- Note the lack of effect in Korean.

Korean dark and light vowels

Korean names

- The most popular baby names in Korea between 2008 and October 27, 2020 as listed at koreanname.me (Kimkkikki, 2020).
- The site draws its data from the national family e-registry.
- Names ranked in the top 1,000 for each gender were extracted from the database.
- Two syllable names are analyzed: n=1,938

Korea names

- About 30% of the name are used for both genders. (e.g., 지우 most popular unisex name)
- 98% of the names are renderable in Sino Korean characters – although not all are likely Sino Korean.
  - Non-Sino Korean names are more common in Female (3.5%) than male names (1.0%).

Sound symbolism in Korean vowels

- Korean sound symbolic words distinguish dark and light vowels, which show vowel harmony.
  - Light vowels are associated with smallness and low intensity.
  - Dark vowels are associated with largeness and high intensity

    toltol vs. tultul $k^hoŋtak$ vs. $k^huntʌk$
  - Conflicts with the height symbolism.

Korean Vowels (Kim-Renaud 1978)

- **Light** = female, Neutral = either, **Dark** = male

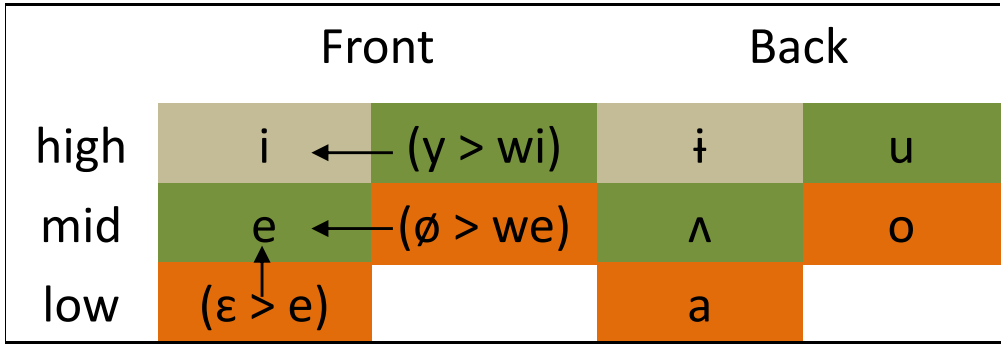
	Front		Back	
high	i	y	ɨ	u
mid	e	ø	ʌ	o
low	ɛ		a	

- **High** = female, Mid = either, **Low** = male

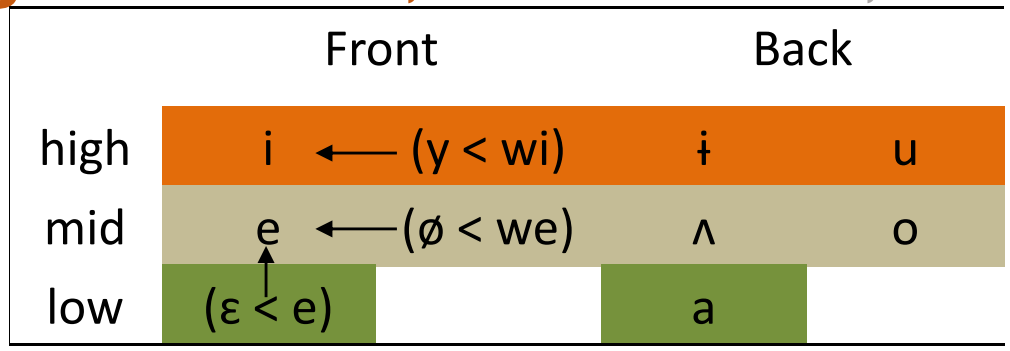
	Front		Back	
high	i	y	ɨ	u
mid	e	ø	ʌ	o
low	ɛ		a	

Korean Vowels: reorganized

- **Light** = female, **Dark** = male, Neutral = either

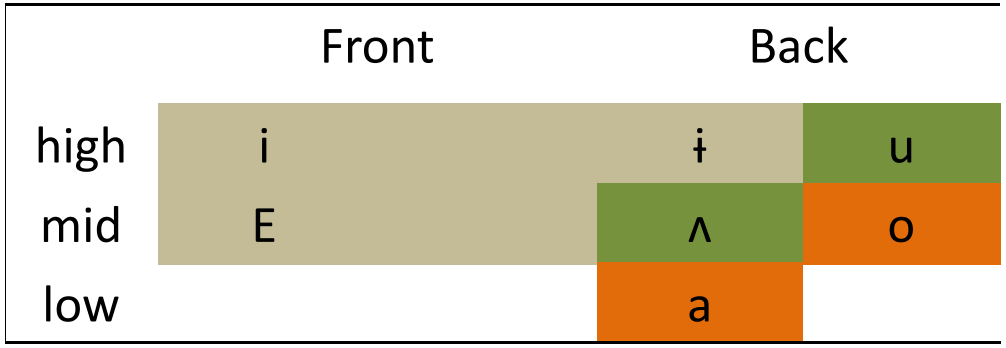


- **High** = female, Mid = either, **Low** = male

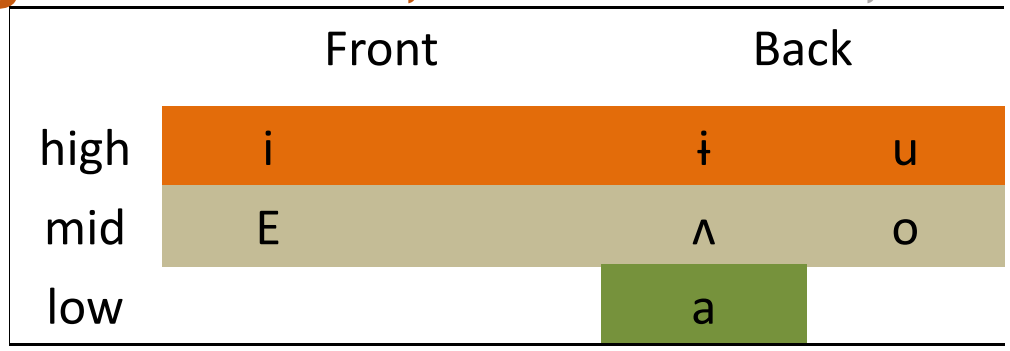


Korean Vowels: reorganized

- **Light** = female, **Dark** = male, Neutral = either

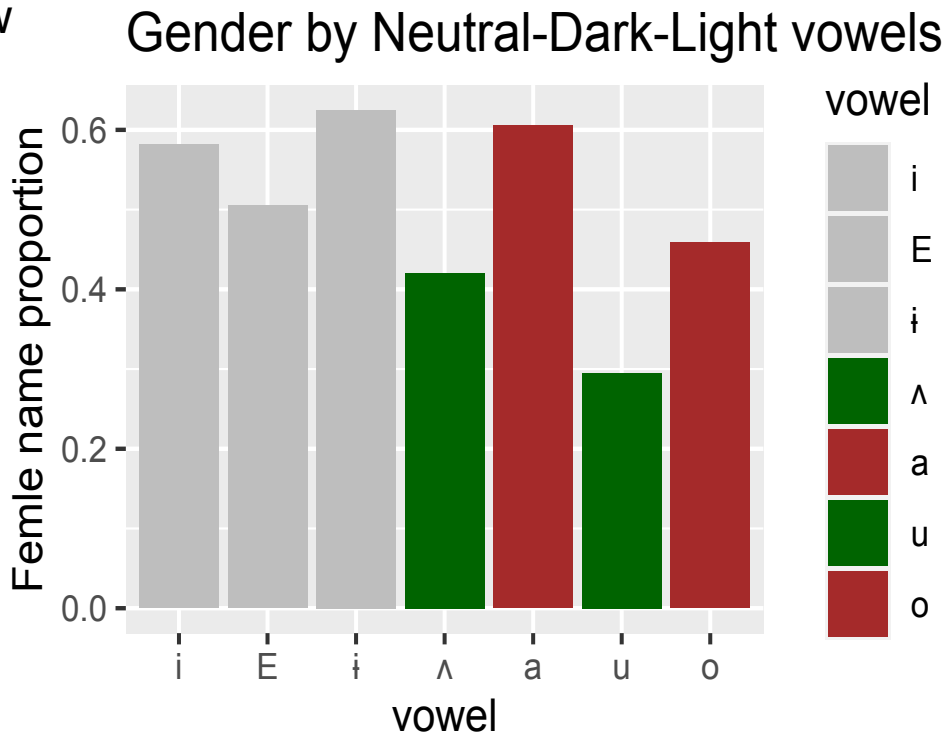


- **High** = female, Mid = either, **Low** = male



Vowel distribution in names

- Native sound symbolism
  - Harmonic dark and light vowels follow the native sound symbolic patterns: Dark – male, Light – female.
- Cross-linguistic sound symbolism
  - Among harmonic vowels, rounded vowels (lower F2 frequency) prefer male names more than unrounded vowels (higher F2 frequency).
  - Neutral vowels, high or front, favor female names.
- Cumulative effects



Maximum Entropy Harmonic Grammar (MaxEnt)

- A stochastic model of OT grammar where constraints are assigned weights, not strict ranking, to capture probabilistic linguistic patterns.
- Constraints contribute to candidate evaluation cumulatively: the weights of the constraints are added up to calculate the probability of candidates.

	A $w=3$	B $w=2$	C $w=1$	$\mathcal{H}$-score	eHarmony	Z	predicted
Candidate 1	1			$1\times3=3$	$e^{-3}=0.0498$	0.0565	88%
Candidate 2		2	1	$2\times2+1\times1=5$	$e^{-5}=0.0067$	0.0565	12%

Constraints

- Default preference
  - *M : don’t be a male
  - *F : don’t be a female
- Korean sound symbolic
  - *Dark – F : a dark vowel must not be in a female name
  - *Light – M : a light vowel must not be in a male name
- Crosslinguistic sound symbolic
  - *Front – M: a front vowel (= higher F2) must not be in a male name
  - *Round – F: a round vowel (= lower F2) must not be in a female name
  - *High – M: a high vowel (= smaller aperture) must not be in a male name
  - *Low – F: a low vowel (= larger aperture) must not be in a female name

OT grammar

		weight	0	0.8975092	0.34508454	0.3034386	0.85028	0.82841737	0.34156466	0.70965943
			Front-F	Round-M	High-F	Low-M	*M	*F	Dark-M	Light-F
		observed	C1	C2	C3	C4	C5	C6	C7	C8
a	M	204.00	0	0	0	0	1	0	0	1
	F	313.00	0	0	0	1	0	1	0	0
E	M	181.00	1	0	0	0	1	0	0	0
	F	185.00	0	0	0	0	0	1	0	0
ʌ	M	201.00	0	0	0	0	1	0	0	0
	F	146.00	0	0	0	0	0	1	1	0
i	M	322.00	1	0	1	0	1	0	0	0
	F	448.00	0	0	0	0	0	1	0	0
ɨ	M	75.00	0	0	1	0	1	0	0	0
	F	125.00	0	0	0	0	0	1	0	0
o	M	170.00	0	0	0	0	1	0	0	1
	F	144.00	0	1	0	0	0	1	0	0
u	M	299.00	0	0	1	0	1	0	0	0
	F	125.00	0	1	0	0	0	1	1	0

MaxEnt constraint weights

- Constraint weights are calculated based on the observed data to maximize the match between the observed probability and the predicted probability.
- Different models of grammar can be tested for their goodness of fit and compared for relative goodness.

			weight	0	0.8975092	0.34508454	0.3034386	0.85028	0.82841737	0.34156466	0.70965943							
				Front-F	Round-M	High-F	Low-M	*M	*F	Dark-M	Light-F							
			observed	C1	C2	C3	C4	C5	C6	C7	C8	harmony	eharmony	z	p	ln(p)		loglike
	a	M	204.00	0	0	0	0	1	0	0	1	1.5599402	0.21014864	0.53258291	0.39458389	-0.9299235		-1966.5776
		F	313.00	0	0	0	1	0	1	0	0	1.13185597	0.32243427	0.53258291	0.60541611	-0.5018393		
	E	M	181.00	1	0	0	0	1	0	0	0	0.85028077	0.42729495	0.86403488	0.49453437	-0.7041386		
		F	185.00	0	0	0	0	0	1	0	0	0.82841737	0.43673994	0.86403488	0.50546563	-0.6822752		
	ʌ	M	201.00	0	0	0	0	1	0	0	0	0.85028077	0.42729495	0.73766746	0.57925145	-0.5460186		
		F	146.00	0	0	0	0	0	1	1	0	1.16998203	0.31037252	0.73766746	0.42074855	-0.8657199		
	i	M	322.00	1	0	1	0	1	0	0	0	1.1953653	0.3025934	0.73933333	0.40927871	-0.8933589		
		F	448.00	0	0	0	0	0	1	0	0	0.82841737	0.43673994	0.73933333	0.59072129	-0.526411		
	ɨ	M	75.00	0	0	1	0	1	0	0	0	1.1953653	0.3025934	0.73933333	0.40927871	-0.8933589		
		F	125.00	0	0	0	0	0	1	0	0	0.82841737	0.43673994	0.73933333	0.59072129	-0.526411		
	o	M	170.00	0	0	0	0	1	0	0	1	1.5599402	0.21014864	0.38815668	0.54140158	-0.613594		
		F	144.00	0	1	0	0	0	1	0	0	1.72592657	0.17800804	0.38815668	0.45859842	-0.7795804		
	u	M	299.00	0	0	1	0	1	0	0	0	1.1953653	0.3025934	0.42909615	0.70518787	-0.349291		
		F	125.00	0	1	0	0	0	1	1	0	2.06749123	0.12650275	0.42909615	0.29481213	-1.221417		

MaxEnt models

- Model comparisons

constraint	weight
*M	0.85
*F	0.83
***High-M**	**0.35**
***Low-F**	**0.3**
***Front-M**	**0**
***Round-F**	**0.9**
***Light-M**	**0.71**
***Dark-F**	**0.34**

constraint	weight
*M	0.88
*F	0.80
***High-M**	**0.27**
***Low-F**	**0**
***Front-M**	**0**
***Round-F**	**0**

constraint	weight
*M	0.85
*F	0.83
***Light-M**	**0.71**
***Dark-F**	**0.34**

LogLikelihood= -1967 LogLikelihood= -2056 LogLikelihood = -2031

Conclusion

Summary

- Some sound ~ gender connections are consistently found across languages and reliably extended in nonce word judgment tasks.
  - Vowel backness, height
  - Consonant sonorancy
  - Syllable weight
- Language-specific “systematic” sound symbolic patterns may partially override certain “iconic” tendencies.
  - Korean dark-light vowels vs. vowel height
- Sound symbolism as part of productive phonological grammar, that interacts with language-specific phonological systems.
- The cumulative and probabilistic patterning can be modeled with the same grammatical tools that model core phonological phenomena.

Thanks to…

- Audience at MOT 2019, ICPhS 2019, Stony Brook and USC.
- Nathan Sanders, Avery Ozburn
- Funding:
  - SSHRC Doctoral Fellowship to Lisa Sullivan
  - University of Toronto Excellence Award to K. Wong
  - Jackman Humanities Institutes Scholars in Residence Program
  - University of Toronto Scarborough Research Competitiveness Fund