Modeling universal and lexical influences on phonotactic judgments

James Myers and Jane Tsay
National Chung Cheng University
Lngmyers *at* ccu.edu.tw
http://www.ccunix.ccu.edu.tw/~lngmyers/

TPC-4 – Taipei, Taiwan – May 4, 2013

Overview

- Universality enhances lexical learning
  – Cooperative interaction
- Evidence from acceptability judgments
  – English
  – Southern Min
  – Mandarin (two separate tests)
- Implications for phonological theory
  – Optimality Theory makes the wrong predictions...
  – ... but there may be an OT-like solution

How do universals shape grammar?

- **Universal Grammar (UG)**
  – Experience arranges universal bits of grammar
  – E.g. learning OT ranking (Tesar & Smolensky 2000)
- **Learning biases**
  – Experience is filtered through universal biases
  – E.g. locality in constraint building (Hayes & Wilson 2008)
- UG and biases make distinct predictions
  – Strict vs. fuzzy universals (e.g. Mielke 2008)
  – **How universals and experience interact** ☜

Clues from acceptability judgments

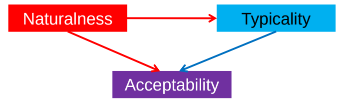
- Native speaker judgments of nonwords reveal productive phonological knowledge
- **Lexical typicality** improves acceptability
  – English-like items sound better to English listeners
  – Likewise for many languages, including Chinese (e.g. Bailey & Hahn 2001; Myers & Tsay 2005)
- **Naturalness** also improves acceptability
  – Universally less marked items sound better (e.g. Frisch & Zawaydeh 2001; Hayes & White 2013)

Operational definitions

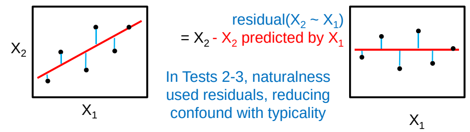
- **Lexical typicality:** Learned from experience
  – Highly analytical (e.g. constraint weights)
  – Sort of analytical (e.g. phonotactic probability)
  – Holistic (e.g. neighborhood density)
- **Naturalness:** Universally helps learning
  – Typology (by hunch, or quantitative)
    - Cross-linguistically common = easier to learn
  – Complexity (e.g. in terms of features)
    - Simpler = easier to learn

Reducing statistical confounds

- Naturalness & lexical typicality are correlated:



- A statistical trick: Replace one independent variable with residuals:

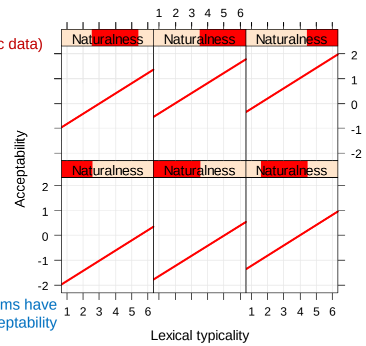


Three logically possible interactions

- **No interaction**
  – Judgments improve with lexical typicality equally strongly in natural and unnatural items
- **Competitive interaction**
  – Judgments show stronger lexical typicality effects in unnatural items
- **Cooperative interaction**
  – Judgments show stronger lexical typicality effects in natural items

No interaction

(Schematic data)



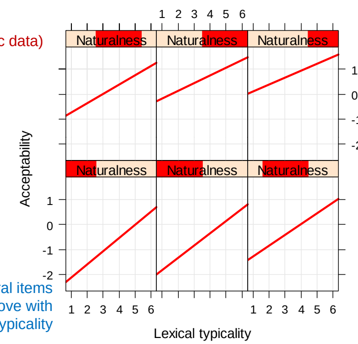
Natural items have higher acceptability

Lexical typicality improves acceptability by the same amount, regardless of naturalness

Unnatural items have lower acceptability

Competitive interaction

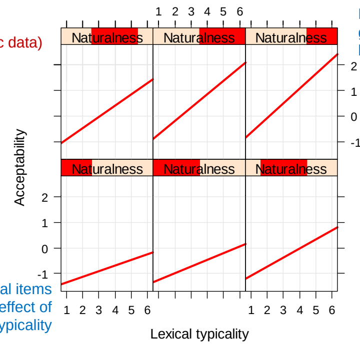
(Schematic data)



Natural items show little effect of lexical typicality

Unnatural items greatly improve with lexical typicality

Cooperative interaction

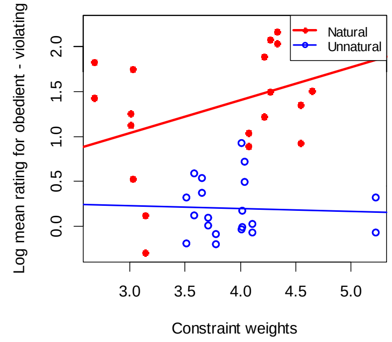


Test 1: English

- Reanalyzing the constraint-violating nonword judgment data from Hayes & White (2013)
- **Lexical typicality:** Weights of computer-learned constraints (higher = less typical)
- **Naturalness:** Typological intuitions (albeit by experts)

		Naturalness of constraints	
		Unnatural	Natural
Constraint weights	Lower	**oid** (vs. oit)	**trefk** (vs. treft)
	Higher	**ooker** (vs. ocker)	**jouy** (vs. jout)

Result: Cooperation

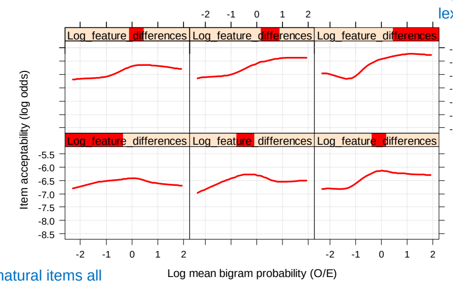


Test 2: Southern Min

- 255 non-lexical syllables, one per each logically possible bigram of Southern Min phonemes
- 20 native speakers, binary auditory judgments
- **Lexical typicality:** Lexical bigram probability (observed / expected: Frisch & Zawaydeh 2001)
- **Naturalness:** Bigram feature differences (more differences = easier to distinguish perceptually; tone was ignored...)

		Mean number of feature changes	
		Lower	Higher
Mean bigram probability (O/E ratio)	Lower	**biem^{1}**	**sot^{4}**
	Higher	**guoŋ^{7}**	**piok^{8}**

Result: Cooperation



Two tests in Mandarin

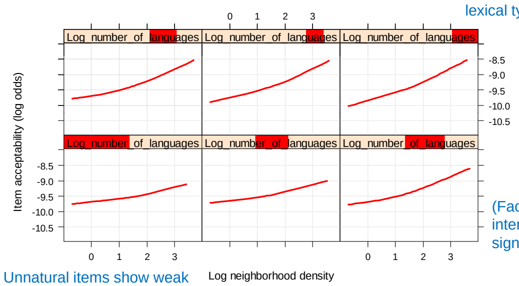
- A **mega-study** (see e.g. Balota et al. 2012)
  - All 3,274 non-lexical syllables that can be written in BPMF (Taiwan's phonetic orthography)
  - Binary judgments of BMPF syllables
- **Test 3:** Pilot (16 speakers)
  - Neighborhood density x onset typological frequency
- **Test 4:** Full mega-study (76 speakers)
  - Phonotactic probability within vs. across languages
- Both analyses ignore tone again

Test 3: Neighbors & Onset typology

- **Lexical typicality:** Number of lexical neighbors
  One segment from target (Vitevitch & Luce 1999)
- **Naturalness:** Number of languages that have the item's initial consonant (in UPSID; Maddieson 1984)

		UPSID frequency of onset	
		Lower	Higher
Neighborhood density	Lower	**t^{h}io^{2}**	**pio^{4}**
	Higher	**t^{h}iei^{1}**	**piei^{1}**

Result: Cooperation



Test 4: Lexical & typological phonotactics

- **Lexical typicality:** Joint transition probability of bigrams in Mandarin (Albright 2009):
  $(freq(s_1 s_2)/freq(s_1)) \times (freq(s_2 s_3)/freq(s_2)) \times ...$
- **Naturalness:** Same thing, but computed over languages in ASJP database (Brown et al. 2013)

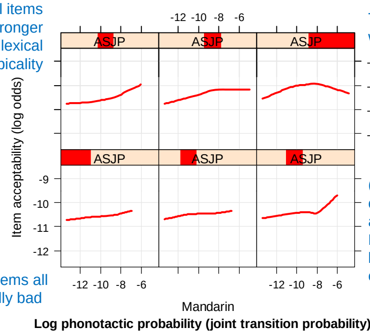
(see Appendix 1 for details on ASJP)

		ASJP transition probability	
		Lower	Higher
Mandarin transition probability	Lower	**siau^{4}**	**sin^{3}**
	Higher	**liai^{1}**	**fau^{2}**

Result: Cooperation?

More natural items show stronger effect of lexical typicality

Then something weird happens...



(Factors and curving interaction are significant: More natural = lexical typicality effect more wiggly!)

Unnatural items all about equally bad

Is the cooperative interaction real?

- Empirically robust?
  - Other quantifications of naturalness and lexical typicality, other languages, interfering factors...
    - Whence the wiggliness...?
- Statistically meaningful?
  - Interaction may be due to a “floor effect”:
    - Are the flatter trend lines merely due to overall low acceptability?
    - But Test 4 shows simple interactions aren’t inevitable

Can OT get this pattern?

- Consider three markedness constraints A, B, C
- Encode items via violation profiles
  - [ABC] = a form obeying all three constraints
  - [aBC] = a form violating A but obeying B & C, etc...
- Sort items by degree of markedness:
  - More natural items: [aBC], [AbC], [ABc]
  - Less natural items: [abC], [aBc], [Abc]
- **Does OT more readily distinguish among the *natural* items, as the judgment data imply?**

OT: No interaction, or competition

- All violation profiles equally distinguishable:

More natural items

	A	B	C
aBC	*		
AbC		*	
ABc			*

Less natural items

	A	B	C
abC	*	*	
aBc	*		*
Abc		*	*

- Unnatural items more informative about constraint demotion:

	C	A	B
☞abC		*	*
aBc	*	*	
Abc	*		*

Interactions and addition

- OT comes from **Harmonic Grammar** (HG), which computes harmony via addition (Hayes & Wilson 2008; Potts et al. 2010):

H = Weight$_1$ x Violations$_1$ + Weight$_2$ x Violations$_2$ + ...

- HG makes the same false predictions as OT
  - If **H([aBC]) > H([AbC])**, then **H([abC]) > H([Abc])**
- In essence this is because interactions go beyond addition (as in statistics)

$$Y = X_1 + X_2 + X_1X_2$$

Modeling interactions in OT

- The X_1X_2 part in $Y = X_1 + X_2 + X_1X_2$ looks familiar...
- Try **constraint conjunction** (Smolensky 1993)
  - A&B violated if and only if both A and B are violated
- But this makes unnatural items *more* distinct:

More natural items

	A&B	A	B	C
aBC		*		
AbC			*	
ABc				*

Less natural items

	A&B	A	B	C
abC	*	*	*	
aBc		*		*
Abc			*	*

Positive constraint conjunction

- Try again (Crowhurst & Hewitt 1997; cf. Wolf 2007)
  - AB is *obeyed* if and only if both A and B are *obeyed*
  - (The Smolensky type is actually disjunction)
- This works! Natural items are more distinct:

More natural items

	AB	A	B	C
aBC	*	*		
AbC	*		*	
ABc				*

Less natural items

	AB	A	B	C
abC	*	*		
aBc	*		*	
Abc	*			*

(See Appendices 2 & 3 for more discussion...)

Sketch of a learning model

- Learner encodes items via innate constraints
  - /san/ = {✓ONSET, ✓*NUC/i, ×NOCODA,...}
  - This is the key **learning bias**
- Learner creates positively conjoined constraints
  - [aBC] triggers creation of BC [abC] triggers nothing
- And ranks them by how often they’re obeyed
  - freq([aBC]) > freq([AbC]) triggers BC >> AC
  - The above steps capture cooperative interaction
- Non-conjoined constraints do nothing...?
  - BC >> AC works (almost) exactly like B >> A

Conclusions

- Naturalness and lexical typicality cooperate
  - Many empirical questions remain open
  - (The ASJP database is an amazing typology tool)
- Standard OT misses this insight
  - Positive conjoined constraints may help
  - Many formal questions remain open
- Learning biases are better than UG
  - UG: Learning fills innate gaps (no interaction, or competition)
  - Biases: Naturalness helps learning (cooperation)

Thanks!

- Experimental participants
- Assistants Guo-Ming Hong, Chiung-Wen Hsu, Zi-Ping Hsu, Yu-Leng Lin, Chia-Wen Lo, Chen-Tsung Yang
- National Science Council (Taiwan) grants NSC 97-2410-H-194-067-MY3, NSC98-2410-H-194-086-MY3, NSC101-2410-H-194-115-MY3
- Commentators Tsung-Ying Chen, Bruce Hayes, Elizabeth Hume, Yu-Leng Lin, and the audiences of SLE 45 (Stockholm, 2012) and the Workshop on Phonological Markedness (Hsinchu, Taiwan, 2012)

Appendix 1: The ASJP database

- The Automated Similarity Judgment Program was created to study diachronic phonology
  - Around 40 Swadesch (1971) words per language
  - Currently contains 5844 languages!
    - Includes creoles, Esperanto, Klingon, no sign language...
  - Transcriptions can be simplified and noisy:
    - Mandarin “blood” [ɕye^{3}]: Sie, Swe, Siueh
- But it’s also great for phonological typology
  - Family and genus information can be used for representative samples (we haven’t tried this yet)

back

Appendix 2: Gradient interactions

The more unnatural, the more constraints needed to distinguish...

		AB	AC	AD	BC	BD	CD	A	B	C	D
More natural	aBCD	*	*	*				*			
	AbCD	*			*	*			*		
	ABcD		*		*		*			*	
	ABCd			*		*	*				*
Semi-natural	abCD	*	*	*	*	*		*	*		
	aBcD	*	*	*	*		*	*		*	
	aBCd	*	*	*		*	*	*			*
	AbcD	*	*		*	*	*		*	*	
	AbCd	*		*	*	*	*		*		*
	ABcd		*	*	*	*	*			*	*
Less natural	abcD	*	*	*	*	*	*	*	*	*	
	abCd	*	*	*	*	*	*	*	*		*
	aBcd	*	*	*	*	*	*	*		*	*
	Abcd	*	*	*	*	*	*		*	*	*

Appendix 3: Think positive

- **Standard OT**
  - Representations encoded via violations only (Golston 1996)
  - Learner learns only what is not already innate
  - **UG approach:** What’s innate is a partial grammar
- **OT with positive constraint conjunction**
  - Representations also encode obeyed constraints
  - Innately sanctioned representations filter learning
  - **Bias approach:** What’s innate is a learning algorithm

back

References (1/4)

Albright, A. (2009). Feature-based generalisation as a source of gradient acceptability. *Phonology, 26* (1), 9-41.

Bailey, T. M., & Hahn, U. (2001). Determinants of wordlikeness: Phonotactics or lexical neighborhoods? *Journal of Memory & Language, 44*, 569-591.

Balota, D. A., Yap, M. J., Hutchison, K.A., & Cortese, M. J. (2012). Megastudies: What do millions (or so) of trials tell us about lexical processing? In J. S. Adelman (Ed). *Visual word recognition, Vol. 1*. Psychology Press.

Brown, C. H., Holman, E. W., & Wichmann, S. (2013). Sound correspondences in the world's languages. *Language, 89* (1), 4-29. ASJP website: http://email.eva.mpg.de/~wichmann/ASJPHomePage.htm

References (2/4)

Crowhurst, M., & Hewitt, M. (1997). Boolean operations and constraint interactions in Optimality Theory. UNC at Chapel Hill & Brandeis ms. [ROA-229]

Frisch, S. A., & Zawaydeh, B. A. (2001). The psychological reality of OCP-Place in Arabic. *Language, 77* (1), 91-106.

Golston, C. (1996). Direct Optimality Theory: Representation as pure markedness. *Language, 72* (4), 713-748.

Hayes, B., & White, J. (2013). Phonological naturalness and phonotactic learning. Linguistic Inquiry, 44 (1), 45-75.

Hayes, B., & Wilson, C. (2008). A Maximum Entropy model of phonotactics and phonotactic learning. *Linguistic Inquiry, 39* (3), 379-440.

References (3/4)

Maddieson, I. (1984). *Patterns of sounds*. Cambridge University Press.

Mielke, J. (2008). *The emergence of distinctive features*. Oxford University Press.

Myers, J., & Tsay, J. (2005). The processing of phonological acceptability judgments. *Proceedings of Symposium on 90-92 NSC Projects* (pp. 26-45). Taipei, Taiwan, May.

Potts, C., Pater, J., Jesney, K., Bhatta, R., & Becker, M. (2010). Harmonic Grammar with linear programming: From linear systems to linguistic typology. *Phonology, 27* (1), 77-117.

Smolensky, Paul (1993). Harmony, markedness, and phonological activity. Talk presented at Rutgers Optimality Workshop I, New Brunswick, NJ. [ROA-87]

References (4/4)

Swadesh, M. (1971). *The origin and diversification of language*. Edited by Joel Sherzer. Chicago: Aldine.

Tesar, B., & Smolensky, P. (2000). *Learnability in Optimality Theory*. MIT Press.

Vitevitch, M. S., & Luce, P. A. (1999). Probabilistic phonotactics and neighborhood activation in spoken word recognition. *Journal of Memory and Language, 40*, 374-408.

Wolf, M. (2007). What constraint connectives should be permitted in OT? In M. Becker (Ed.), *University of Massachusetts occasional papers in linguistics 36: Papers in theoretical and computational phonology* (pp. 151-179). Amherst, MA: GLSA.