

## The Origins of Lexical Biases: Evidence from Navajo Compounds and English Naming Preferences

Andy Martin (UCLA)  
amartin@humnet.ucla.edu

### Main claim:

- ❖ Selective word formation plays a role in establishing/maintaining phonotactic lexical biases.

### Two kinds of phonotactic:

- Categorical: some structures are banned outright (e.g., \*[waŋ] in English)
- Gradient: some structures are merely dispreferred; realized as statistical biases in the lexicon

### Example of gradient phonotactic:

Place OCP in Arabic: consonants with the same place of articulation do not tend to occur together in roots, even when nonadjacent (Frisch, Pierrehumbert and Broe 2004)

This cooccurrence restriction is gradient, since violations are permitted, but rare.

### Where do such biases come from?

### Two possible answers:

- Historical change, possibly error-driven, in individual lexical items (Ohala 1981)
- Selective formation/retention of lexical items

### Previous evidence for selective word formation:

- Children's avoidance of words with difficult sounds (Ferguson and Farwell 1975, Leonard et al. 1981, Schwartz and Leonard 1982)

### I. Navajo compound formation

- Sibilants in Navajo compounds tend to agree in anteriority
- This sibilant agreement bias influences compound formation

### Sibilant harmony in Navajo: the basic facts

#### (1) Navajo sibilant classes

[+anterior]: [s, z, ts<sup>h</sup>, ts, ts']

[-anterior]: [ʃ, ʒ, tʃ<sup>h</sup>, tʃ, tʃ']

#### (2) Sibilant harmony is exceptionless when...

- occurring between a sibilant in a root and one in an affix

(a) /ji-s-lé:ʒ/ <sup>1</sup> → [ji-ʃ-tlé:ʒ] 'it was painted'

(b) /ji-s-tiz/ → [ji-s-tiz] 'it was spun'

(Fountain 1998)

- occurring between two sibilants in the same root (roots underlined)

(c) [tʃ'oh-ts'ózi] 'rush grass'

(d) [naʔní-ʒo:ʒ] 'bridge'

*unattested*: roots like \*/tʃ'ózi/ or \*/-zo:ʒ/

(Young, Morgan and Midgette 1992)

When both sibilants are in different roots (i.e. in compounds), however, disagreeing sibilants are repaired only in some words:

#### (3) Compounds

##### Repaired

(Sapir and Hojier 1967)

(a) [tsa:-né:z] 'mule' (/tʃa:/ 'ear' + /né:z/ 'long')

(b) [xos-ts'ó:z] 'type of cactus' (/xoʃ/ 'cactus' + /ts'ó:z/ 'slender')

##### Not repaired

(Young, Morgan and Midgette 1992)

(c) [tʃéi-ts'i:n] 'thorax, chest bone' (/tʃéi/ 'heart' + /ts'i:n/ 'bone')

(d) [ts<sup>h</sup>é-tʃé:ʔ] 'amber' (/ts<sup>h</sup>é/ 'stone' + /tʃé:/ 'resin')

(e) [na:-niʃ-ts<sup>h</sup>oh] 'big job' (/niʃ/ 'work' + /ts<sup>h</sup>oh/ 'big')

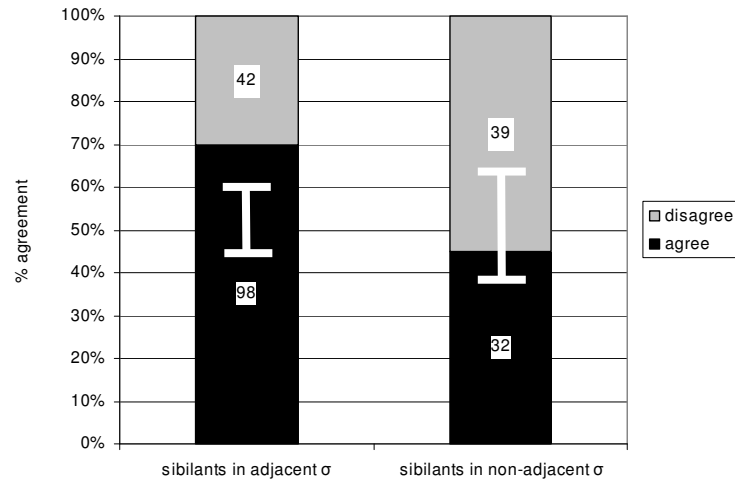
- Harmony is more likely to apply when the sibilants are close together than when they are widely separated (Sapir and Hojier 1967, Martin 2004).

*Question*: Are compounds biased towards sibilant agreement?

Data: All compounds from a Navajo dictionary (Young, Morgan and Midgette 1992) made up of two roots, each of which contains one sibilant (total 211)

<sup>1</sup> Navajo examples are given in IPA, with accents marking high tones.

(4) Anteriority agreement rate: surface only

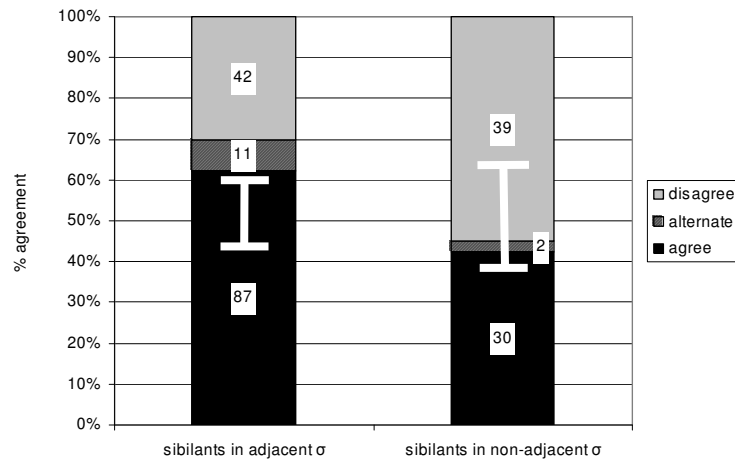


Chance calculated using a Monte Carlo test (Kessler 2001)  
I-shaped bars = 95% confidence interval

Apparent gradient phonotactic: disagreeing sibilants possible, but dispreferred

*Question:* What's causing the elevated agreement rates?  
Is it just that repairs are occurring more frequently at closer distances?

(5) Anteriority agreement rate: surface vs. underlying



- Sibilants are more likely to agree underlyingly when sibilants are closer

Navajo speakers seem to prefer forming (or retaining) compounds whose roots *already* contain agreeing sibilants.

- ❖ **Phonological constraints can influence which words are adopted into the lexicon.**

**II. American first names**

*OCP in English:* words containing two identical liquids (like *rare*) are underrepresented in the lexicon, compared to words with two different liquids (like *lair*) (Frisch 1997, Berkley 2000). Particularly strong in the same syllable, weaker in adjacent syllables.

First names

*Question:* does this gradient phonotactic influence how people choose names?

Data source: top-1,000 first name lists by decade according to U.S. Social Security Administration (available at <http://www.ssa.gov/OACT/babynames/>)

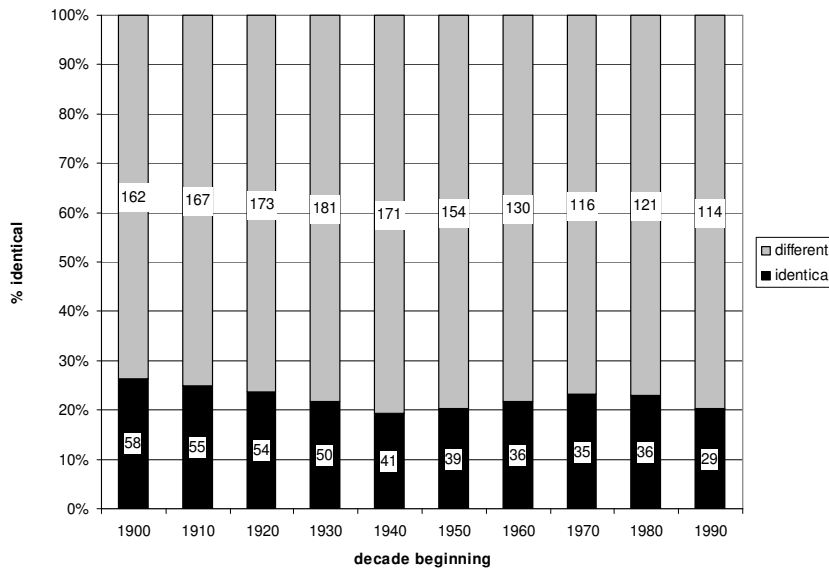
Procedure: for each decade, extract all names containing exactly two liquids, removing cases where the liquids are string-adjacent.

(6) Examples of two-liquid names

different liquids	Claire, Eleanor, Clark, Franklin
identical liquids	Margaret, Delilah, Gertrude, Wilhelm

- How often are the liquids identical? (chance  $\approx$  50%)

(7) Liquid agreement in first names by decade



- OCP effect: all identity rates are below chance (p<.0001 by Monte Carlo)
- No overall trend: no pairwise difference between decades is significant
- Distance effect (stronger OCP in same syllables) is present in all decades (not shown here)

Not merely an effect of list similarity: only 32.3% of the names on the 1900s list occur on the 1990s list.

❖ **Bias in name choice is consistent despite large changes in the actual names involved**

**III. Formal account**

Three possible responses to disagreeing sibilants in Navajo:

- *Toleration* (allow disagreement on the surface)
- *Repair* (fix one of the sibilants)
- *Avoidance* (don't use the word at all)

Proposal:

❖ **Avoidance is the result of low-ranked M-PARSE (Prince and Smolensky 1993)**

(8) Avoidance

/sV+ǰV/	FAITH	AGREE	M-PARSE
(a) sV.ǰV		*!	
(b) ǰV.ǰV	*!		
(c) <i>null</i>			*

Tripartite behavior of Navajo compounds is the result of variable ranking:

(9) Possible rankings

- (a) FAITH, M-PARSE » AGREE *toleration*
- (b) AGREE, M-PARSE » FAITH *repair*
- (c) FAITH, AGREE » M-PARSE *avoidance*

- Different rankings apply probabilistically to different words

**IV. Conclusions**

*Navajo*

- A constraint on sibilant agreement biases formation of compounds

*English*

- A constraint on liquid cooccurrence biases choice of first names

**General conclusions**

- Word-formation is biased—it is constrained not only by absolute prohibitions but by statistical tendencies
- Selective word formation can result in lexical biases for which a purely error-driven account is implausible
- Further evidence for the role of constraints in phonology
- Variable ranking of M-PARSE with respect to markedness and faithfulness constraints can account for the grammar's effect on word formation

**References**

- Berkley, Deborah. 2000. Gradient OCP Effects. Ph.D. dissertation. Northwestern University.
- Burke, Sean and Jordan Lachler. 2000. Online analytical lexicon of Navajo. Web-based version of Young, Morgan, and Midgette 1992. <http://www.speech.cs.cmu.edu/egads/navajo/>
- Ferguson, C.A. and C. Farwell. 1975. Words and sounds in early language acquisition: English initial consonants in the first 50 words. *Language* 51: 419-439.
- Fountain, Amy. 1998. An Optimality Theoretic approach to Navajo prefixal syllables. Doctoral Dissertation. University of Arizona. ROA-238.
- Frisch, Stefan. 1997. Similarity and Frequency in Phonology. Ph.D. dissertation. Northwestern University.
- Frisch, Stefan, Janet Pierrehumbert and Michael Broe. 2004. Similarity Avoidance and the OCP. *NLLT* 22: 179-228.
- Kessler, Brett. 2001. *The Significance of Word Lists*. Stanford: CSLI Publications.
- Leonard, L.B., R.G. Schwartz, B. Morris and K. Chapman. 1981. Factors influencing early lexical acquisition: lexical orientation and phonological composition. *Child Development* 52: 882-887.
- Martin, Andrew. 2004. Gradient distance effects in Navajo sibilant harmony. Talk handout, LSA annual meeting.
- Ohala, John. 1981. The listener as a source of sound change. In C.S. Masek, R.A. Hendrick, and M.F. Miller (eds.) *Papers from the Parasession on Language and Behavior*. Chicago: Chicago Linguistic Society. 178-203.
- Prince, Alan and Paul Smolensky. 1993. Optimality Theory: Constraint Interaction in Generative Grammar. Ms., Rutgers University and University of Colorado, Boulder. ROA-537.
- Sapir, Edward and Harry Hoijer. 1967. *The phonology and morphology of the Navaho language*: University of California Publications in Linguistics 40. Berkeley: University of California Press.
- Schwartz, R.G. and L.B. Leonard. 1982. Do children pick and choose? An examination of phonological selection and avoidance in early lexical acquisition. *Journal of Child Language* 9: 319-336.
- Young, Robert. W., William Morgan Sr. and Sally Midgette. 1992. Analytical Lexicon of Navajo. Albuquerque: University of New Mexico Press.