Featural vs. arbitrary categories

Elliott Moreton
UNC-Chapel Hill

(1) Preview:

a. Topic is the putative analytic bias which has received the most attention in the lab: phonetically-systematic vs. phonetically-arbitrary classes.

b. Typological phenomena it might explain.

c. Review of the empirical studies that have been done on it.

d. What can we conclude? What needs to be done?

1 Phonological rectification

(2) Phonetic patterns are not a direct, unmediated basis for phonology. In particular, phonology is more symmetrical than phonetics — level sets in phonetic difficulty space are not the boundaries between phonologically permitted and forbidden areas. The boundaries get straightened out by the phonological feature system (Hayes, 1999).

(3) Inventories tend to make maximal, or at least more-than-chance, use of feature combinations, avoiding isolated segments whose features aren't shared by others (Clements, 2003). Let [V] represent any voiced labial fricative, such as [v] or [β], and let [P B F] be defined likewise. Then [V] occurs more often in inventories that also have [P B F] than it would if its occurrence did not depend on theirs:

	Has [P B F]		≥ 1 missing		
	Act.	Exp.	Act.	Exp.	Total
has [V]	83	51	64	96	147
lacks [V]	74	106	230	198	304
Total	157		294		451

(4) Difficulty of producing voiced vs. voiceless stops is influenced by many factors: PoA, closure duration, whether postnasal, position in phrase. You don't get constraints like "*>25 effort units" (defining effort w.r.t. a particular aerodynamic model), which would translate into, e.g., "*bdg after obstruents, *dg initially, *g after oral sonorants". Instead, you get "* voiced obstruent after another obstruent", "*voiced velar obstruent", etc. — symmetrical regions of phonological space, not regions bounded by level sets in phonetic space (Hayes, 1999).

(Phonetic difficulty is a kind of channel bias. If something is "difficult", it is difficult <u>to do correctly</u>, which means it is likely to be done <u>in</u>correctly, thus introducing a systematic bias into the production-perception channel.)

(5) ⇒ Perhaps there is an analytic bias in favor of phonetically-systematic classes as contrasted with phonetically-arbitrary ones.

(6) This is <u>not</u> the same thing as saying that phonetically-unsystematic classes are unlearnable in natural language or in the lab.

a. They are widespread in natural languages (Mielke, 2004, Ch. 4). Some of them are unproductive, but then, plenty of phonetically-systematic patterns are unproductive too (Zimmer, 1969; Hsieh, 1976; Moreton, 2002; Zhang et al., 2006; Becker et al., 2007).

b. They can be learned to the same level of performance as systematic ones in the lab (Seidl and Buckley, 2005; Peperkamp and Dupoux, 2007)

A typologically-effective analytic bias could be subtle, as could a typologically-effective channel bias.

---

2 Experimental evidence

(7) Basic question: If a task requires learning to respond differently to two sets of speech sounds, is it easier when the sets are distinguished by a phonetic feature than when they are phonetically arbitrary?

⇒ Not interested in studies that contrast one arbitrary pattern with another, like the ones that are meant to simulate lexical learning (Saffran et al., 1996,?; Peña et al., 2002; Onishi et al., 2002; Chambers et al., 2003), etc.

(8) We are concerned here only with the question of whether a class is easier to learn if it is defined by a phonetic property, regardless of whether that property is an official [feature] in some particular theory of features (which is a related, but separate, question).

Terminology: "Featural" vs. "Arbitrary". (I'll avoid "natural", which has too many meanings.)

2.1 Category learning, no morphology or alternations

2.1.1 Categories are segments

(9) Saffran and Thiessen (2003): 9-month-old infants, L1 English. 3-phase experiment:

a. Familiarize on 60-word nonsense lists conforming to a particular pattern, A or B (2 minutes).

b. Play continuous stream of speech made from 4 new words, 2 of type A and 2 of type B (1 minute).

c. Test on each of the 4 words from (b). Use Headturn Preference Procedure, with stimulus consisting of repetitions of a single word. Dependent measure is total looking time. 3 (pseudo-)replications of each trial.

(10) Rationale: Familiarization teaches phonotactic pattern, which facilitates segmentation, which then leads to a difference in looking times in the test phase.

I.e., (b) and (c) form a typical segmentation experiment, and the idea is that preceding it with (a) will help in segmenting those words that fit the pattern. (Nine-month-olds known to use L1 phonotactics in segmentation (Mattys and Jusczyk, 2001).)

However, not clear whether Phase (b) of exp. actually did anything; their results could have been a direct effect of Phase (a) on Phase (c).

(11) Taxing on the babies: To get 30 usable subjects, had to run 52 (Exp. 1), 59 (Exp. 2), 59 (Exp 3)—almost half fussed out or were otherwise lost.

(12) Experiment 2: voicing/aspiration.

a. CVCCVC, where CVC = [ptk]V[bdg] in the "−+" condition, [bdg]V[ptk] in the "+−" condition. Made by flipping each word in one condition 180 degrees.

b. Extra potential non-statistical cue: longer vowels before /b d g/ than /p t k/ , by about 14ms. No oneÕs found that such a small difference in vowel duration can affect infants in any previous study, so they ignore it.

c. Results: Novel pattern, 95% CI = 7.05s ± 0.72s; familiar pattern, 6.05s ± 0.76s. "Significant" difference.

d. ⇒ 9mos can learn relatively fine-grained phonotactic rules, at least to the extent of distinguishing stimuli with /p t k/ in a particular position from those with /b d g/ in that position.

(13) The pattern in Exp. 2 is typologically bizarre, but we don't know that learners actually induced that pattern. They could have been attending only to the initial consonant, only to the the final consonant, only to the presence of aspiration, etc., or they could have been learning a list of which phonemes occurred in which position, without abstracting a class.

(14) Experiment 3: arbitrary classes

a. Like Exp 2., only /p t k/ vs. /b d g/ was replaced with /p d k/ vs. /b t g/ , so that the two sets could no longer be distinguished by a single feature.

b. Results: SEs were the same size as before, but now there was no difference in means (about 7.5s for both familiar and novel test items).

c. ⇒ In Exp. 3, they must have been abstracting a class rather than just learning a list of phonemes/position pairs.

(15) LaRiviere et al. (1974): Supervised category learning by adult English L1 speakers.

a. Heard *Ca* syllables and learned to sort them into two categories by immediate feedback.

b. Two groups of participants got the same syllables divided up differently: Either the categories were featurally separable, or they were arbitrary.

c. 10 people in each group for each set of syllables.

d. Mixed results (not all shown):

Condition			Results		
Feature	Syllables		Sep.	Diff.?	Arb.
[cont]	[va tʃa za] vs.	[pa da ka]	0.71	=	0.64
	[ʒa ha ʃa va] vs.	[pa ba da ka]	0.68	=	0.60
[strid]	[va tʃa za] vs.	[ra na θa]	0.67	=	0.66
	[ʃa dʒa za sa] vs.	[ða la ka ha]	0.79	≠	0.55
[nas]	[am an aŋ] vs.	[að ab adʒ]	0.81	≠	0.58
[voice]	[ða ga za] vs.	[pa sa ʃa]	0.73	=	0.68
	[da za va ba] vs.	[pa ka ʃa θa]	0.62	=	0.57
	[ba da ga] vs.	[pa ta ka]	0.55	=	0.55

e. Why such weak results for [voice], compared to Saffran and Thiessen (2003)?

(16) Cristiá and Seidl (2008): Unsupervised category learning by 7-month-old L1 English learners.

a. Exp. 1. 24 7mos. 60 C_1VC_2 pseudowords. Language condition determined C_1: m, n, and two obstruents.

"Natural" condition (nasals & stops, [–cont])

Phase	Version A				Version B			
Familiarization	m	n	t	g	m	n	b	k
Test			b	k			t	g

"Arbitrary" condition (nasals & fricatives)

Phase	Version A				Version B			
Familiarization	m	n	v	ʃ	m	n	f	z
Test			f	z			v	ʃ

b. 60 pseudowords in each condition. **H**: "[When] in the natural condition a given word began with a stop, in the arbitrary condition, the corresponding word began with a fricative."

c. Familiarize on 57 pseudowords in 19 "sentences" of 3 each, played once. Test on 6 trials, 3 stop-initial and 3 fricative-initial, using onsets that had not appeared in initial position during familiarization. Headturn Preference Procedure, except that familiarization was not contingent on looking.

d. Results: 10/12 looked longer to illegal in Natural cond., whereas only 5/12 did in Arb. The mean looking times differed reliably also:

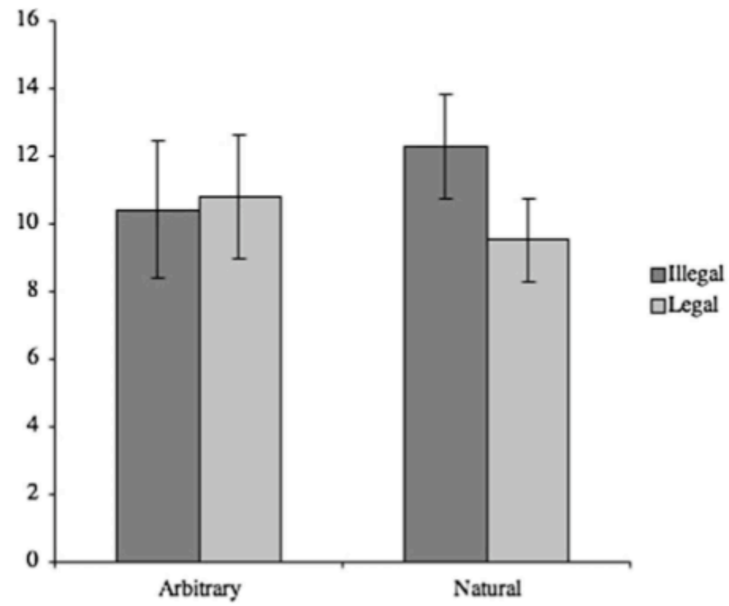


FIGURE 1 Looking times by Condition and Legality in Experiment 1 (error bars represent standard error).

e. Exp. 2. Could that be just because it's hard to process fricatives, or learn any constraint involving them? Same as Exp. 1, except that the familiarization phase omitted the nasal-initial words. If the difficulty in Exp. 1 Arb came from grouping fricatives with nasals, then removing the nasals should remove the difficulty, whereas if it came from processing fricatives, then it should stay. Results: Main effect of legality, no interaction with Condition.

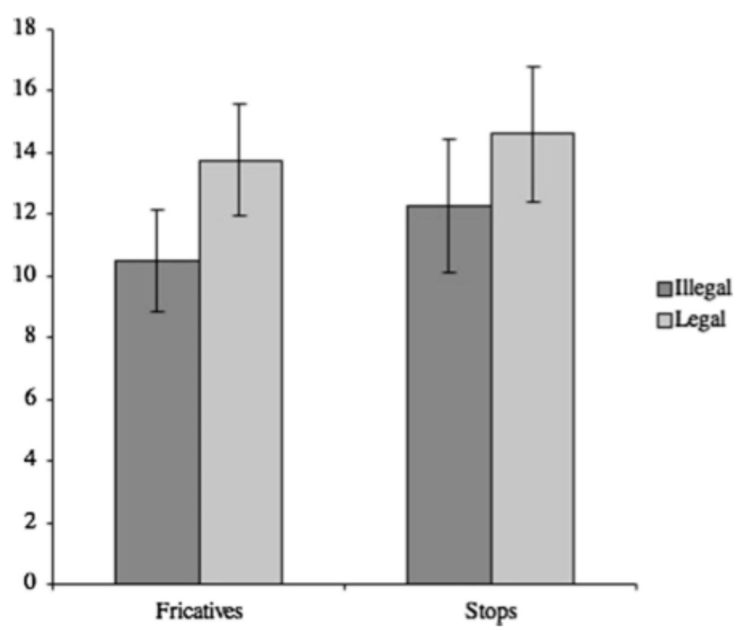


FIGURE 3 Looking times by Condition and Legality in Experiment 2 (error bars represent standard error).

f. Oddly, this time they got longer looks to legals than illegals from 17/24 participants. You often get a novelty preference in an easy task and familiarity preference in a difficult one, but this seems to be the reverse. C&S suggest that the classes in Exp. 2 were more complex than those in Exp. 1, since they require more features to specify.

2.1.2 Categories are clusters (phonotactics)

(17) Kuo (2009): Acquisition of new onset phonotactics by Mandarin speakers. Mandarin Chinese allows syllables of the form Onset-Glide-Vowel-(X). There are L1 restrictions on onset-glide combinations (Duanmu, 2003, 27–29):

	Labial		Dental		Retroflex		Palatal		Velar	
Cj	pj	p^{h}j	tj	t^{h}j	--	--	tɕj	tɕ^{h}j	--	--
Cw	(pwɔ)	(p^{h}wɔ)	tw	t^{h}w	tʂw	tʂ^{h}w	--	--	kw	k^{h}w

(18) Three phonotactic dependency conditions. Two versions of each, swapping legal/illegal to check whether prior bias made one or the other easier (it didn't in any of the cases).

a. "Language P[lace]": Glide is conditioned by place of articulation.

Version A

	Labial		Dental	
Cj	~~pj~~	~~p^{h}j~~	tj	t^{h}j
Cw	(pwɔ)	(p^{h}wɔ)	~~tw~~	~~t^{h}w~~

Version B

	Labial		Dental	
Cj	pj	p^{h}j	~~tj~~	~~t^{h}j~~
Cw	~~(pwɔ)~~	~~(p^{h}wɔ)~~	tw	t^{h}w

b. "Language L[aryngeal]": Glide is conditioned by aspirated/unaspirated.

Version A

	Labial		Dental	
Cj	pj	~~p^h~~ j	tj	~~t^{h}j~~
Cw	~~(pwɔ)~~	(p^{h}wɔ)	~~tw~~	t^{h}w

Version B

	Labial		Dental	
Cj	~~pj~~	p^{h}j	~~tj~~	t^{h}j
Cw	(pwɔ)	~~(p^{h}wɔ)~~	tw	~~t^{h}w~~

c. "Language N[either]": Glide is conditioned by PoA and aspiration jointly:

Version A

	Labial		Dental	
Cj	pj	~~p^{h}j~~	~~tj~~	t^{h}j
Cw	~~(pwɔ)~~	(p^{h}wɔ)	tw	~~t^{h}w~~

Version B

	Labial		Dental	
Cj	~~pj~~	p^{h}j	tj	~~t^{h}j~~
Cw	(pwɔ)	~~(p^{h}wɔ)~~	~~tw~~	t^{h}w

(19) Experimental procedure:

Study 1	1:20	Listen to stimuli (24 disyllabic reduplicated nonsense words $\times$ 1 repetition).
Distraction 1	0:40	Arithmetic problems.
Test 1	4:00	2AFC, which sounds more like language? One legal, one not. 10 legals are old (heard in Study 1), 10 are new (legal onsets with new rimes).
Distraction 2	0:40	Arithmetic problems.
Study 2	4:00	Listen to stimuli again (24 $\times$ 3 repetitions).
Distraction 3	0:40	Arithmetic problems.
Test 2	8:00	Test 1 again twice in different order.

(20) Participants: 30–31 per condition, half in each version. From vocational school in Taiwan, 16–18 years old. All L1 Mandarin. Some exposure to Southern Min. One more was run, but dropped because exit interview showed she explicitly figured out the pattern in the PoA condition.

(21) Results (pooled across test blocks): Proportion correct, significance by *t*-test vs. chance (=0.50):

Condition	Old		New	
PoA	0.68	***	0.62	***
Aspiration	0.67	***	0.59	***
Neither	0.58	*	0.51	n.s.

No significant difference between Place and Laryngeal conditions, despite within-language PoA/glide relationships. Both Place and Laryngeal were significantly better than Neither.

2.2 Learning allomorph selection

(22) Pycha et al. (2003): Adult L1 English speakers learning vowel harmony (backness).

a. Stimulus design:

Stem			
	i	u	
C_1	ɪ	ʊ	C_2
	æ	a	

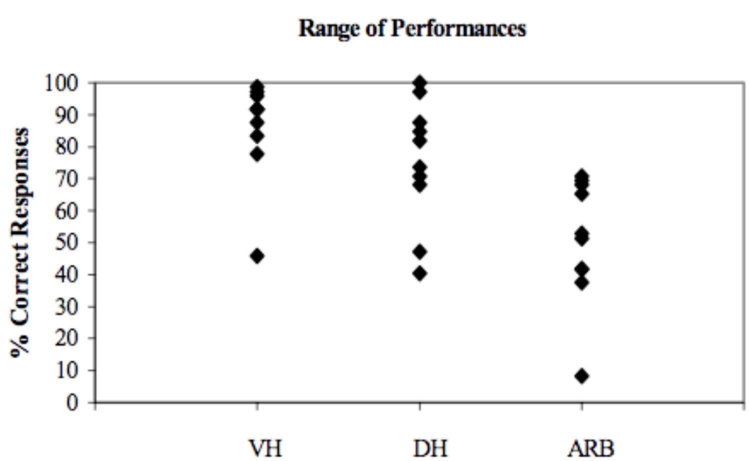
Suffix (“plural”)			
	Condition		
Suffix	Harmony	Disharmony	Arbitrary
–ɛk	i ɪ æ	u ʊ a	i æ ʊ
–ʌk	u ʊ a	i ɪ æ	ɪ a u

b. Training: Listen to 18 “singular/plural” pairs twice. Then, listen to 36 pairs (old and new), repeated twice. Half were pattern-conforming, the other half not. Press button for correct/incorrect, get feedback. (Participants never said anything out loud.)

c. Test: Like training, but the 36 tokens were totally new, and there was no feedback.

d. 10 participants in each condition; L1 English. No previous exposure to vowel-harmony language.

e. Results: Harmony ≈ Disharmony > Arbitrary.



(23) We've already seen Pycha et al. (2007), which found that English L1 speakers acquired both progressive and regressive place-assimilation patterns better than an assimilation pattern which was progressive for some places of articulation and regressive for others:

Condition	Familiarization Stimulus	Familiarization Response	Test Stimulus	Test Response	% conforming
Regressive	kut	kukkut	pat	?	95
	deb	deddeb	gub	?	
Progressive	kut	kuttut	pat	?	93
	deb	debbeb	gub	?	
Arbitrary	kut	kukkut	pat	?	47
	deb	debbeb	gub	?	

(24) Peperkamp et al. (2006): Featural vs. arbitrary consonant alternations, acquired by L1 French adults.

a. Participants are familiarized on Det+N phrases paired with pictures of two or three objects, where Det is one of [nel] = 2 or [ra] = 3. The initial C of the N may undergo alternations:

Condition	Underlying	Alternation
Featural A	ptk bdg fsʃ	/f s ʃ/ ⟶ [v z ʒ] /V _ V
Featural B	ptk fsʃ vzʒ	/p t k/ ⟶ [b d g] /V _ V
Arbitrary A	p k bdg sʃ v ʒ	/p g z/ ⟶ [ʒ f t] /V _ V
Arbitrary B	pt dg f ʃ vzʒ	/ʃ v d/ ⟶ [b k s] /V _ V

b. Training: Listen to 20 phrases, each accompanied by picture. Nouns beginning with non-dentals appear with both nel and ra. (Those beginning with dentals appeared with only one of the two, so that they could be used to test generalization of the alternation to new segments. There wasn't any, and I'll ignore dentals hereafter.)

c. Test: Two parts. People in both Featural conditions got the same test items; those in both Arbitrary conditions did too.

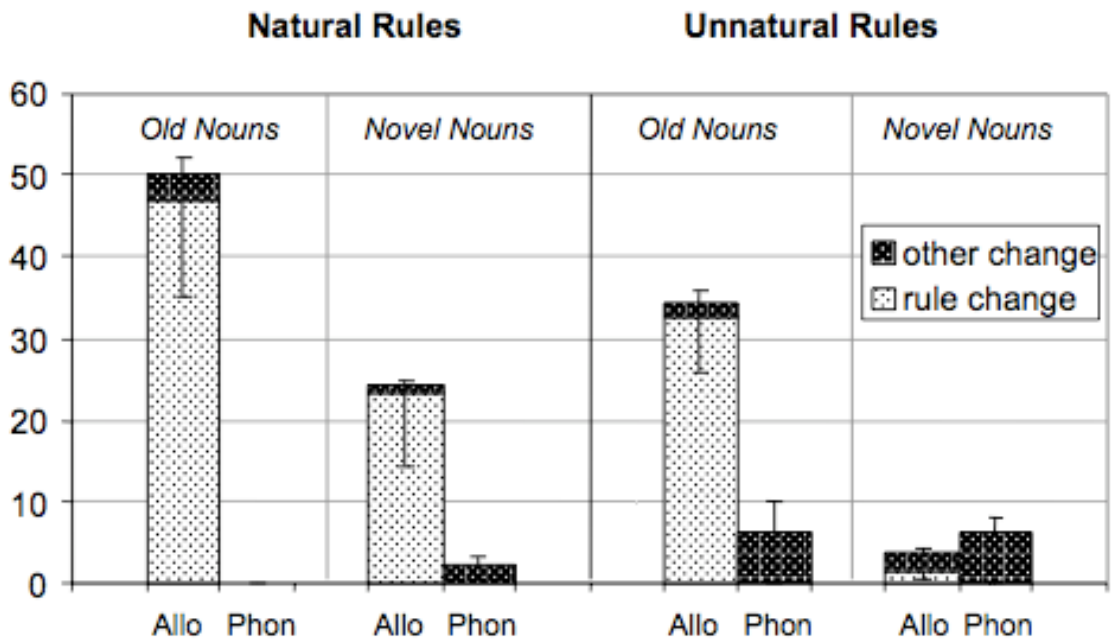
(i) 12 test trials for 12 old nouns from training phase.

(ii) 48 test trials for new nouns, half stop-initial, half fricative-initial

Test works like this: Get 2-sec exposure of a phrase-picture pair that was identical in the A and B version. Then, get a picture with a different number of the same object, and produce appropriate phrase.

d. 32 L1 French speakers, 8 in each group.

e. Results:



("Phonemic" condition means C shouldn't change, while "Allophonic" means it should. "Natural/Unnatural" means "Featural/Arbitrary".)

f. ⇒ The Featural condition elicited better performance on old nouns, and more transfer to new ones.

(25) The same authors, with the same stimuli but a different task (phrase-picture matching, rather than production) found no difference between featurally-systematic and featurally-arbitrary patterns (Peperkamp and Dupoux, 2007):

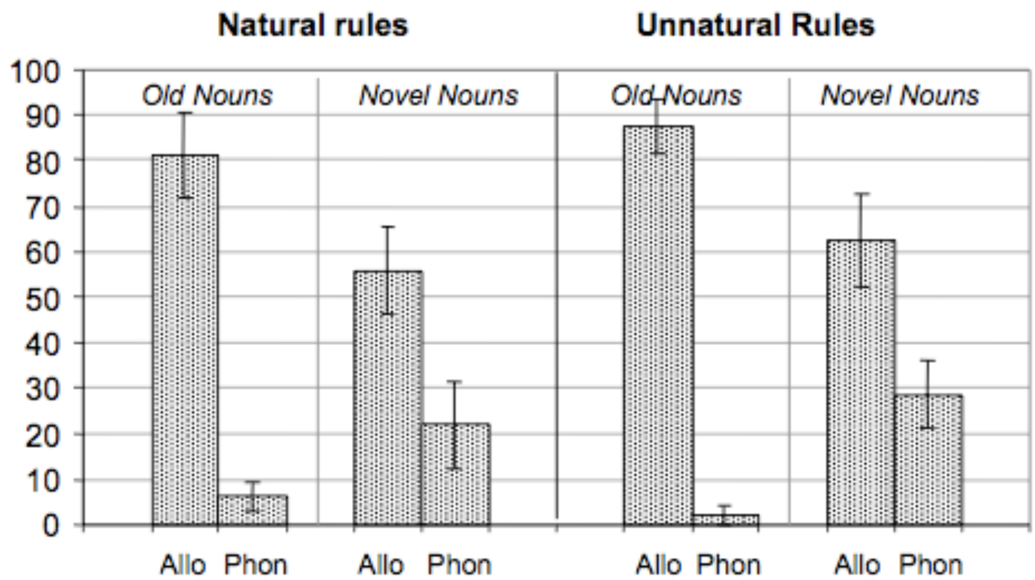


Figure 1. Mean percentages of matching of a phrase presented with a change in the initial obstruent of the noun to an image representing the unchanged noun. Data from P&D.

They suggest that the phrase-picture matching task is easier, leaving more resources over for learning.

(26) Healy and Levitt (1980, Exp. 3): Acquisition of an analogue of the English -*s* rule by L1 English speakers.

a. English cluster phonotactics cause underlying - /z/ and - /d/ suffixes to devoice after a voiceless C (dogs/cats, dog's/cat's, stands/sits, raised/clamped, armored/helmeted, etc.). Will that facilitate acquisition of an analogous rule in an artificial language?

b. Stimuli were VC syllables (all combinations of 10 vowels by /p t k b d g/). Participants were told they were learning to attach a gender suffix to a root in an invented language (i.e., instructions led them to believe that the predictive dependency was lexical, not phonological).

c. 3 affix conditions: s/z, f/v, and a/o. Crossed with 2 trigger conditions: featural (ptk goes with s/f/a) and arbitrary (ptg goes with s/f/a).

d. 60 trials, consisting of: Tape pronounces the VC syllable; 6 sec. silence, during which participant circles one of two letters on answer sheet; tape says, "The correct answer is:"; tape names the letter that should follow; participant then pronounces the resulting VCC or VCV aloud.[1] No training/test distinction. Dependent measure was error rate across whole experiment.

e. Results: Featural better than arbitrary in the two conditions where the pattern was voicing assimilation, but no difference in the one where allomorph selection was phonetically unrelated to the conditioning environment:

[1]There was a "silent" condition as well, which got similar results. I'm omitting it for simplicity's sake.

Stem-final C	Affix -s/-z	-f/-v	-a/-o
ptk vs. bdg	0.19	0.24	0.45
ptg vs. bdk	0.38	0.43	0.44
Different?	Yes	Yes	No

Errors in the Arbitrary condition tended to be on Vk or Vg stimuli, suggesting that people were grouping them with their voice-mates.

(27) What can we conclude from this? It is certainly consistent with the hypothesis that featurally-systematic patterns are easier to learn, but the English confound looms large:

a. English L1 voicing assimilation rule could prime the -s/-z and -f/-v conditions, explaining better performance there than on -a/o.

b. Likewise, English priming could also explain the difference between the ptk/bdg and ptg/bdk conditions.

3 Discussion

(28) Featural vs. arbitrary has been studied in the lab more intensively than any other putative analytic bias. How solid is the evidence for it?

(29) What experimental factors tend to bring out the differences?

(30) The existence of a bias in the lab is no guarantee that it has any effect on typology. It might be too weak; it might be overwhelmed by channel bias; there might be some other analytic bias in the opposite direction that the experiment wasn't set up to detect, etc.

Worse yet, featural-vs.-arbitrary bias is inextricably confounded with channel bias.

What kind of experiment (broadly construed) could we conceivably do that could provide positive evidence that featural-vs.-arbitrary bias is typologically effective?

(31) What kind of learning model would exhibit a "soft" bias against featurally-arbitrary classes (slower learning, but not absolute impossibility)?

Might it have anything to do with category-learning models in other cognitive domains (Billman and Knutson, 1996; Ashby et al., 1999; Love et al., 2004; Feldman, 2006)? If so, what?

References

Ashby, F. G., S. Queller, and P. Berretty (1999). On the dominance of unidimensional rules in unsupervised categorization. Perception and Psychophysics 61(6), 1178–1199.

Becker, M., N. Ketrez, and A. Nevins (2007). Where and why to ignore lexical patterns in Turkish obstruent alternations. Handout, 81st annual meeting of the Linguistic Society of America, Anaheim, California.

Billman, D. and J. Knutson (1996). Unsupervised concept learning and value systematicity: a complex whole aids learning the parts. Journal of Experimental Psychology: Learning, Memory, and Cognition 22(2), 458–475.

Chambers, K. E., K. H. Onishi, and C. Fisher (2003). Infants learn phonotactic regularities from brief auditory experience. Cognition 87, B69–B77.

Clements, G. N. (2003). Feature economy in sound systems. Phonology 20(3), 287–333.

Cristiá, A. and A. Seidl (2008). Is infants' learning of sound patterns constrained by phonological features? Language Learning and Development 4(3), 203–227.

Duanmu, S. (2003). The phonology of Standard Chinese (1st ed.). Oxford, England: Oxford University Press.

Feldman, J. (2006). An algebra of human concept learning. Journal of mathematical psychology 50, 339–368.

Hayes, B. (1999). Phonetically driven phonology: the role of optimality in inductive grounding. In M. Darnell, E. Moravcsik, M. Noonan, F. Newmeyer, and K. Wheatly (Eds.), Functionalism and Formalism in Linguistics, Volume 1: General Papers, pp. 243–285. Amsterdam: John Benjamins.

Healy, A. F. and A. G. Levitt (1980). Accessibility of the voicing distinction for learning phonological rules. Memory and Cognition 8(2), 107–114.

Hsieh, H. (1976). On the unreality of some phonological rules. Lingua 38, 1–19.

Kuo, L. (2009). The role of natural class features in the acquisition of phonotactic regularities. Journal of psycholinguistic research 38(2), 129–150.

LaRiviere, C., H. Winitz, J. Reeds, and E. Herriman (1974). The conceptual reality of selected distinctive features. Journal of Speech and Hearing Research 17(1), 122–133.

Love, B. C., D. L. Medin, and T. M. Gureckis (2004). SUSTAIN: a network model of category learning. Psychological Review 111(2), 309–332.

Mattys, S. L. and P. W. Jusczyk (2001). Phonotactic cues for segmentation of fleunt speech by infants. Cognition 78(2), 91–121.

Mielke, J. (2004). The emergence of distinctive features. Ph. D. thesis, Ohio State University.

Moreton, E. (2002). Structural constraints in the perception of english stop-sonorant clusters. Cognition 84, 55–71.

Onishi, K. H., K. E. Chambers, and C. Fisher (2002). Learning phonotactic constraints from brief auditory experience. Cognition 83, B13–B23.

Peña, M., L. L. Bonatti, M. Nespor, and J. Mehler (2002). Signal-driven computations in speech processing. Science 298, 604–607.

Peperkamp, S. and E. Dupoux (2007). Learning the mapping from surface to underlying representations in an artificial language. In J. S. Cole and J. I. Hualde (Eds.), Papers in Laboratory Phonology 9, pp. 315–338.

Peperkamp, S., K. Skoruppa, and E. Dupoux (2006). The role of phonetic naturalness in phonological rule acquisition. In D. Bamman, T. Magnitskaia, and C. Zoller (Eds.), Papers from the 30th Boston University Conference on Language Development (BUCLD 30), pp. 464–475.

Pycha, A., P. Nowak, E. Shin, and R. Shosted (2003). Phonological rule-learing and its implications for a theory of vowel harmony. In M. Tsujimura and G. Garding (Eds.), Proceedings of the 22nd West Coast Conference on Formal Linguistics (WCCFL 22), pp. 101–114.

Pycha, A., E. Shin, and R. Shosted (2007). Directionality of assimilation in consonant clusters: an experimental approach. MS, Department of Linguistics, University of California, Berkeley.

Saffran, J. R., R. N. Aslin, and E. L. Newport (1996). Statistical learning by 8-month-old infants. Science 274, 1926–1928.

Saffran, J. R., E. L. Newport, and R. N. Aslin (1996). Word segmentation: the role of distributional cues. Journal of Memory and Language 35(4), 606–621.

Saffran, J. R. and E. D. Thiessen (2003). Pattern induction by infant language learners. Developmental Psychology 39(3), 484–494.

Seidl, A. and E. Buckley (2005). On the learning of arbitrary phonological rules. Language Learning and Development 1(3 & 4), 289–316.

Zhang, J., Y. Lai, and C. Turnbull-Sailor (2006). Wug-testing the "tone circle" in Taiwanese. In D. Baumer, D. Montero, and M. Scanlon (Eds.), Proceedings of the 25th meeting of the West Coast Conference on Formal Linguistics (WCCFL), Somerville, Mass., pp. 453–461. Cascadilla Proceedings Project.

Zimmer, K. E. (1969). Psychological correlates of some Turkish morpheme structure constraints. Language 45(2), 309–321.