24.941J / 6.543J / 9.587J / HST.727J The Lexicon and Its Features
Spring 2007

Distinctive features in phonology
part 1: learning

24.941/6.976

Functions of DF's

- **Guide learning and generalization beyond observed data**

- Define natural classes

- Structure and compress the phone inventory

- Define representations in mental lexicon and in mapping lexical entries to sounds perceived and produced

Learning rules

-s after p,t,k,f,θ	-z after b,d,g,v,ð	-z after s, z, ʃ, ʒ, tʃ, dʒ
book[s]	bag[z]	ri[tʃəz]
text[s]	bed[z]	le[dʒəz]
lap[s]	lab[z]	lea[səz]
buff[s]	lo[vz]	lea[ʃəz]
ba[θs]	soo[ðz]	buz[əz]

Segment based rules

- $\varnothing \rightarrow \partial$ {**tʃ, dʒ, ʃ, s, z**}_ **z**
- z -> s/{**p, t, k, f, θ**}_
- Learning each set:

  for each segment, you wait to get positive evidence that it is a member of the set

- Or else you define the set through properties that all and only its members share:

- Ø ->∂ /[+strident]_**z**

- z -> s/[-voice]_

- Then you might not need positive evidence for the behavior of each member

Albright and Hayes (Cognition 2003)

- "Minimal generalization" learner of alternations
- Take each learning pair as a word specific rule

  ➔ lab   lab[z]

- Structural description and structural change

  ➔ Ø -> [z]/ [ɪæb_]ₚₗ

- Compare rules

  ➔ kid   kid[z]

- Find narrowest rule that covers both cases, using a feature description to collapse different SD's

  ➔**[+syll, -back,-round] [-son,+voic,-cont]**

   ɪ   **æ**   **b**

   k   **l**   **d**

Albright and Hayes (Cognition 2003)

- "Minimal generalization" learner of alternations
- Take each learning pair as a word specific rule

    ➔ lab   lab[z]

- Structural description and structural change

    ➔ Ø -> [z]/ [læb_]$_{pl}$

- Compare rules
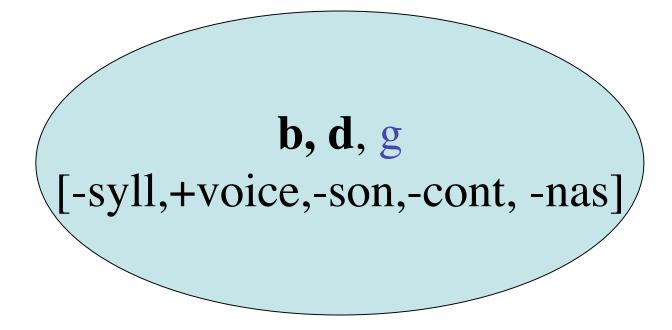
    ➔ kid   kid[z]
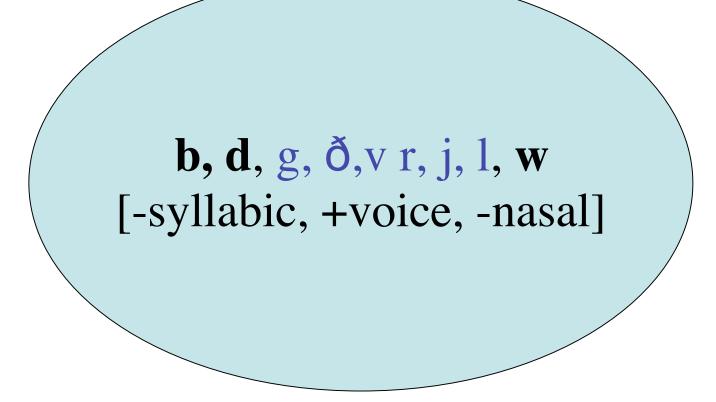
- Find narrowest rule that covers both cases; use a feature description to collapse different SD's

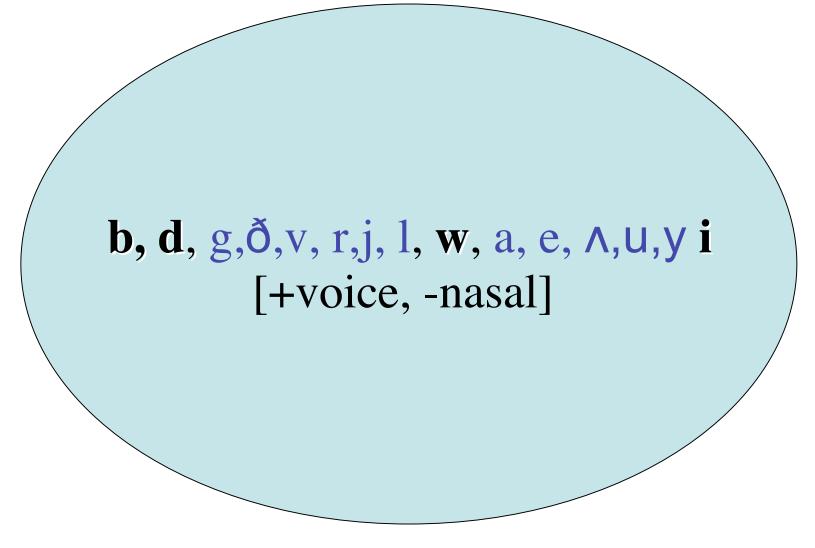   ➔ Ø -> [z]/C [+syll, -back,-round] [-son,+voic,-cont]/_

Spotty evidence

- Will force the learner to generalize beyond the observed data
- Subject to the limits imposed by his feature theory

b, d, g
[-syll,+voice,-son,-cont, -nas]

b, d, g, ð,v r, j, l, **w**
[-syllabic, +voice, -nasal]

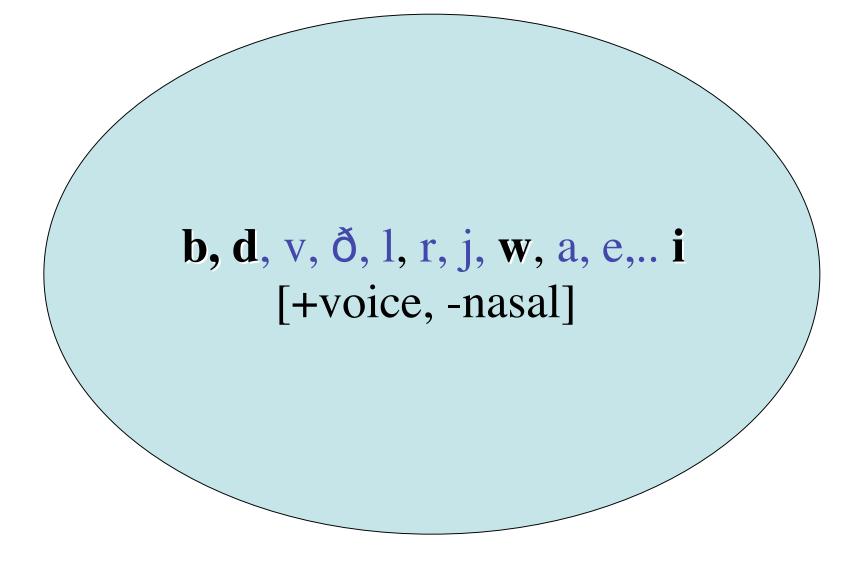**b, d**, g,ð,v, r,j, l, **w**, a, e, ʌ,u,y **i**
[+voice, -nasal]

Process is feature-dependent

- the learner will generalize differently if she operates with SPE's place feature theory

	b	d	g
anterior	+	+	-
coronal	-	+	-

b, d, v, ð, l, r, j, **w**, a, e,.. **i**
[+voice, -nasal]

Anecdotal evidence

- Ba[x], Ba[xs], out-Ba[xt]

- bei[ʒ], bei[ʒəz]

- Suggests

a. that speakers form featurally defined categories, to which novel phones are automatically assigned, without any evidence other than their sound quality

b. the process learned depends on assignment of distinctive feature values.

14

Checking two points

1. How many rules:

   feature analysis permits some descriptions, but not others, to be unified as one rule: is the difference in complexity between analyses reflected in the learning process?

2. Forced generalization:

   feature analysis forces even a conservative learner to make predictions about segments not yet observed. Is this confirmed? (Also: is the learner really conservative?)

Pycha et al. 2003: AG experiment; task is judgment of correctness

- *Palatal Vowel Harmony* (VH):
  -ɛk: CiC-ɛk, CɪC-ɛk, CæC-ɛk
  -ʌk: CuC-ʌk, CʊC-ʌk, CaC-ʌk

- *Palatal Vowel Disharmony* (DH):
  -ʌk: CiC-ʌk, CɪC-ʌk, CæC-ʌk
  -ɛk: CuC-ɛk, CʊC-ɛk, CaC-ɛk

- *Palatal Arbitrary* (ARB):
  After [i, æ, ʊ], front suffix -ɛk: CiC-ɛk, CæC-ɛk, CʊC-ɛk
  After [ɪ, u, a], back suffix -ʌk: CɪC-ʌk, CuC-ʌk, CaC-ʌk

Target rules: segments

- Harmony:

  ʌ-> ɛ / {i, ɪ, æ }C__

- Disharmony

  ʌ-> ɛ / {u, ʊ, a }C__

- Arbitrary

  ʌ-> ɛ -> / {ɪ, u, a}C__

Target rules: features

- Harmony:

  ʌ-> ɛ / [-back, +syllabic] C__
- Disharmony

  ʌ-> ɛ / [+back, +syllabic] C__
- Arbitrary

  ʌ-> ɛ / [+high, +tense, +back, +syllabic]C__

  ʌ-> ɛ / [+high, -tense, -back, +syllabic]C__

  ʌ-> ɛ / [+low, +back, +syllabic]C__

Courtesy of Anne Pycha. Used with permission. Source: Pycha, A., P. Nowak, E. Shin, and R. Shosted. "Phonological Rule-learning and its Implications for a Theory of Vowel Harmony." Edited by G. Garding, M. Tsujimura. Proceedings of the West Coast Conference on Formal Linguistics (WCCFL) 22. (pp. 423-435). Somerville, MA: Cascadilla Press.

Range of Performances

Courtesy of Anne Pycha. Used with permission. Source: Pycha, A., P. Nowak, E. Shin, and R. Shosted. "Phonological Rule-learning and its Implications for a Theory of Vowel Harmony." Edited by G. Garding, M. Tsujimura. Proceedings of the West Coast Conference on Formal Linguistics (WCCFL) 22. (pp. 423-435). Somerville, MA: Cascadilla Press.

16. Harmony: Distribution of errors

CiC-ɛk	CɪC-ɛk	CæC-ɛk	CuC-ʌk	CʊC-ʌk	CaC-ʌk
12	16	12	12	8	38

17. Disharmony: Distribution of errors

CiC-ʌk	CɪC-ʌk	CæC-ʌk	CuC-ɛk	CʊC-ɛk	CaC-ɛk
28	23	36	20	28	44

18. Arbitrary: Distribution of errors

CiC-ɛk	CɪC-ʌk	CæC-ɛk	CuC-ʌk	CʊC-ɛk	CaC-ʌk
46	57	35	57	92	78

Errors in the Arb condition

- 92 on CʊC-ɛk:
- Maybe a rule is partly learned that disfavors [ɛ] after back vowels:
- ʌ -> ɛ /[–back, +tense]C_

Do these results say anything about the role of features in learning rules?

- 2 differences betw. Harmony/Disharmony vs. Arb
  - Complexity of the analysis (how many rules, symbols)
  - Relevance of context to the nature of the change.
- Even if the set {ɪ, u, a} forms a class describable by some feature set, it is probably irrelevant to the change in backness.
- Wilson (2003) shows that equally complex, feature-based rules are learned differently in an AG experiment:
  - Nasal Harmony: **dome-na, suto-la, doke-la**
  - Nasal Disharmony: **dome-la, suto-na, doke-na**
  - Random(velar triggered nasalization): **dome-la, suto-la, doke-na**
  Random differs in having a context unrelated to the change.

Wilson 2003, comparison of dissimilation and random conditions

(8) Mean (SD) proportion "yes" by group and item type

		Group 2A		Group 2B	
		Grammaticality		Grammaticality	
		gramm.	ungramm.	gramm.	ungramm.
Stem	old	.73 (.17)	.47 (.17)	.68 (.12)	.52 (.19)
Type	new	.50 (.13)	.35 (.12)	.47 (.15)	.41 (.15)

Figure by MIT OpenCourseWare. Data from Wilson, Colin. "Experimental Investigation of Phonological Naturalness."
Edited by G. Garding and M. Tsujimura. Proceedings of the West Coast Conference on Formal Linguistics (WCCFL) 22.
Somerville, MA: Cascadilla Press, 2003.

Back to Pycha et al.

- The results don't speak clearly for a difference in ease of learning that's due to feature-based complexity.

- As in Wilson's case, the Arbitrary rule could be hard to learn because its context is unrelated to the change.

- Still unclear: does the feature analysis clarify the relation of the context to the change?

Saffran and Thiessen 2003
Exp. 2: infants; task is segmentation

- 30 9 mo olds; DVT vs. TVD conditions
- Stage 1

  pattern induction: hear a list of CVCCVC conforming to a DVT/TVD template

- Stage 2

  segmentation: hear stream containing repeated old and new CVCCVC; new CVCCVC vary in whether they conform to earlier template

- Difference in listening times for streams of new words that fit the template vs. streams that don't.
- Novelty vs. familiarity preference: here it was novelty.

Sample stimuli

TVD	DVT
todkad	dakdot
kigpid	dipgik
kobtig	gitbok
pudkad	gakdip

Target patterns

- In segment terms

  {p, t, k}_ {b, d, g};

  {b, d, g} _ {p, t, k}

- In feature terms:

  [+voice, -son]_[-voice, -son]

  [-voice, -son]_ [+voice, -son]

Target patterns

- In segment terms
  {p, t, k}_ {b, d, g};
  {b, d, g} _ {p, t, k}

- In feature terms:
  [+voice, -son]_[-voice, -son]
  [-voice, -son]_ [+voice, -son]

Exp 3

- {p, d, k} _ {b, t, g} condition
- {b, t, g} _ {p, d, k} condition
- Induction and segmentation phases are the same as in Exp. 2.

Target patterns

- In segment terms: as complex as pattern of Exp 2.

  a. {p, d, k} _ {b, t, g}

  b. {b, t, g} _ {p, d, k}

- In feature terms: much more complex than Exp2. Here is the feature translation of (a):

$$\left\{ \begin{array}{l} \text{[-voice, -cor]} \\ \text{[+voice, -son, +cor]} \end{array} \right\} \underline{\quad} \left\{ \begin{array}{l} \text{[+voice, -son, -cor]} \\ \text{[-voice, +cor]} \end{array} \right\}$$

Will infants learn to tell the difference between (a) and (b) with equal ease in this case?

Figure 1. Listening times to test words consistent (familiar template) and inconsistent (novel template) with the phonological structures heard during the pattern induction phase of Experiments 1, 2, and 3.

Saffran, J. R., and E. D. Thiessen. "Pattern Induction by Infant Language Learners."

Maybe this result is not relevant either

- There *is* a difference in learning the DVT/TVD pattern of Exp2 vs. the random pattern of Exp3.

- But is it due to infants' use of [±voice] as a classificatory property of segments?

- Or to the fact that the syllable templates in Exp.2 - but not on Exp. 3 - can be described in terms of a global amplitude contour?

Testing three points

1. How many rules:

   feature analysis permits some descriptions, but not others, to be unified as one rule

2. **Forced generalization:**

   feature analysis forces even a conservative learner to make predictions about segments not yet observed.

Finley & Badecker 2007; AG Task: which word is in the language

Poverty of Stimulus Method (Wilson 2006):

➔ In the training phase, withhold from the stimulus set a class of segments.

➔ In the testing phase, expose subjects to the full set of segments.

➔ Observe if pattern learned in training is extended to the set withheld in training.

Extrapolation

	a	b	c
		$\alpha F, \gamma H$	
	γH		
training		b	c
testing	a	b	c

Extrapolation

	æ	e	i
		−low, +syllabic	
	+syllabic		
training		e	i
testing	æ	e	i

38

Interpolation

	a	b	c
		$\alpha F, \gamma H$	
	$\beta G, \gamma H$		
	γH		
training	a		c
testing	a	b	c

Interpolation

	æ	e	i
		−low, +syllabic	
	−high, +syllabic		
	+syllabic		
training	æ		i
testing	æ	e	i

Exp 1: back harmony in high V

- Stem: CVCV
  C from {p, t, k, b, d, g, m, n}
  V from {i, u, e, o, æ, a}
- Suffix: -mi/-mu
- Target rule: [+high] -> [aback]/ [aback] C_
- 3 Conditions
  Low Hold-Out: stem vowel {i, e, u, o}

  Mid Hold-Out: stem vowel {i, æ, u, a}

  Control: stems only, harmonic and disharmonic.
- Testing: forced choice, which word is in the language?

Exp 1, more detail

- Training: 24 pairs {stem, stem+harmonic suffix}; 5x each
  Mid Hold-Out: **buda, buda-mu; bæni, bæni-mi,**
  Low Hold-Out: **budo, budo-mu; bide, bide-mi**
  Control: 24 harmonic, 24 disharmonic stems: **muku, bigu**
- Task: Hear 2 suffixed forms (e.g. **bidi-mu, bidi-mi**}
  Choose which is "most likely to be in the language".
- Test items:
  Old items (heard in training): e.g. **bugu-mu**
  New items (identical V's as in training): e.g. **tuku-mu**
  New vowels (the held-out V's) in these stems:
     Mid Hold-Out: **nike-m**i; Low Hold-Out: **nuka-mu**

Figure 1: Proportion of Harmonic Responses for All Conditions: Experiment 1

Interpretation: Low Hold-Out

- We observe no extrapolation

	æ	e	i
		−low, +syllabic	
	+syllabic		
training		e	i
testing	æ	e	i

Interpretation: Low Hold-Out

- Possible cause: the learner is conservative

	æ	e	i
		–low, +syllabic	
		+syllabic	
training		e	i
testing	æ	e	i

Interpretation: Low Hold-Out

- Alternative: low triggers are disfavored (if high targets).

		e	i
		−low, +syllabic	
		+syllabic	
training		e	i
testing	æ	e	i

Interpretation: Mid Hold-Out

- We observe an interpolation in this condition

	æ	e	i
		−low, +syllabic	
	+syllabic		
training	æ		i
testing	æ	e	i

Interpretation: Mid Hold-Out

Possible cause: the feature set lacks of a feature grouping {low, high}. This forces subjects to learn a broad rule, including mid vowels in the trigger set.

	æ	e	i
		−low, +syllabic	
	+syllabic		
training	æ		i
testing	æ	e	i

Interpretation: Mid Hold-Out

Alternative: Low triggers are disfavored even with overt
evidence. There may be a hierarchy of triggers, where
low V's are at the bottom. The learner knows the
hierarchy and infers: if [æ] is a trigger, [e] should be
one too.

	æ	< e	< i
training	æ		i
testing	æ	e	i

Interpretation: Mid Hold-Out

- Low triggers are disfavored even with overt evidence.
- Hierarchy: if [æ] is a trigger, [e] should be one too.
- Independent evidence favors this.
- Unclear if we need assumptions about the feature set

	æ	< e	< i
training	æ		i
testing	æ	e	i

Exp. 2

- Low suffix: mak/mæk.
- 2 Conditions: stem V in the training set

  Mid Hold-Out: stem vowel   {i, æ, u, a}

  High Hold-Out: stem vowel {e, æ, o, a}

- All else is the same

Figure 2: Proportion of Harmonic Responses for All Conditions: Experiment 2

Interpretation: Mid-Hold-Out

• A feature-based interpolation effect.
• Low triggers ok in this case: perhaps bec. targets are low too.

	æ	e	i
		−low, +syllabic	
	+syllabic		
training	æ		i
testing	æ	e	i

Interpretation: High-Hold-Out

• We observe an extrapolation effect, casting some doubt on the conservative learner hypothesis

	æ	e	i
	−high, +syllabic		
	+syllabic		
training	æ	e	
testing	æ	e	i

Interpretation: High-Hold-Out

- High triggers not so ok: perhaps bec. targets are low.
- Lends support to the idea that target-trigger similarity plays a role

	æ	e	i
	−high, +syllabic		
	+syllabic		
training	æ	e	
testing	æ	e	i

Overall

- Exp. 1 and 2 suggest the following:

  ➔ Bias for similar trigger-target pairs (cf. Rose&Walker)

      explains bad performance on lowV in Exp1; and MHO-HHO in Exp.2

  ➔ Learning is feature- not segment-based.

      explains high overall performance relative to Control in Exp.2

  Not clear we need this:

  ➔ Learner is moderately conservative:

  more willing to interpolate than extrapolate.

Finley and Badecker

- appeal to slightly different biases:

  ➜ Bias against low triggers (Exp.1)

  ➜ Bias against i-triggers (Exp.2)

- The typological or other basis of this is unclear.

- Low triggers are ok in Exp. 2 and High triggers are ok in Exp 1.

Exp. 3: Height Harmony

- Non-low suffix: mi/me.
- 2 Conditions: stem V in the training set

➔ Lax Hold-Out: stem vowel {i, u, e, o}

➔ Back Hold-Out: stem vowel {i, ɪ, e, ɛ}

Lax-Hold-Out: training

	mid	high
tense	e o	i u

Lax-Hold-Out: testing

	mid	high
tense	e o	i u
lax	ɛ ɔ	ɪ ʊ

Back-Hold-Out: training

	mid	high
front	e ɛ	i ɪ

Back-Hold-Out: testing

	mid	high
front	e ɛ	i ɪ
back	o ɔ	u ʊ

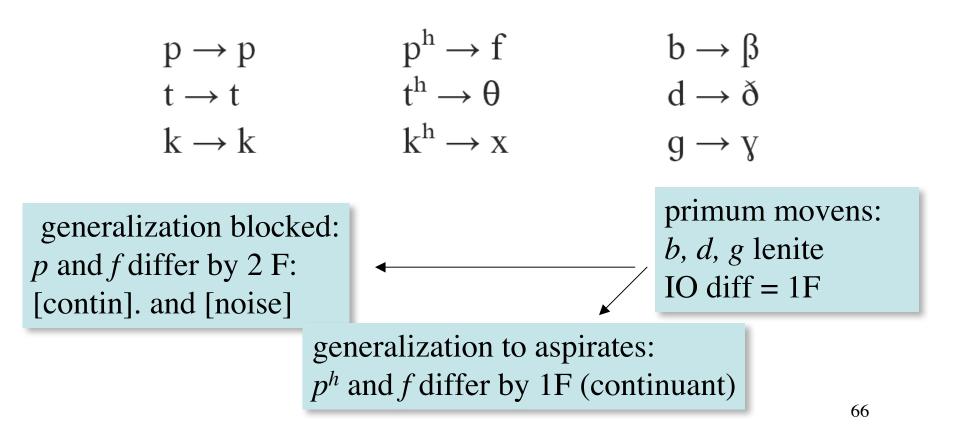Figure 3: Experiment 3 Data

Interpretation

- The learner is not conservative, but biased.
- Extrapolates in Back-Hold-Out condition, from [-back, -low, +syllabic] to [(-low), +syllabic].
- Nature of the bias? "backness can't affect height"

  [perhaps a problem here: no preference for similar triggers-targets]

- Does not extrapolate in the Lax-Hold-Out condition.
- Bias? Lax high is less high than tense high. So [ɪ, ʊ] might be disfavored as raising triggers.

  [but [ɛ, ɔ] might be favored as triggers of lowering]

64

Graff 2006

- Learner is not conservative, but constrained by the assumption that all instances of a process will exhibit a constant I-O distance, where the distance is measured in features.

- Next slides come from Graff's paper.

Greek (1th -4th cent AD)

$$p \rightarrow p \qquad p^h \rightarrow f \qquad b \rightarrow \beta$$
$$t \rightarrow t \qquad t^h \rightarrow \theta \qquad d \rightarrow \eth$$
$$k \rightarrow k \qquad k^h \rightarrow x \qquad g \rightarrow \gamma$$

generalization blocked:
p and *f* differ by 2 F:
[contin]. and [noise]

primum movens:
b, d, g lenite
IO diff = 1F

generalization to aspirates:
p^h and *f* differ by 1F (continuant)

Experiment

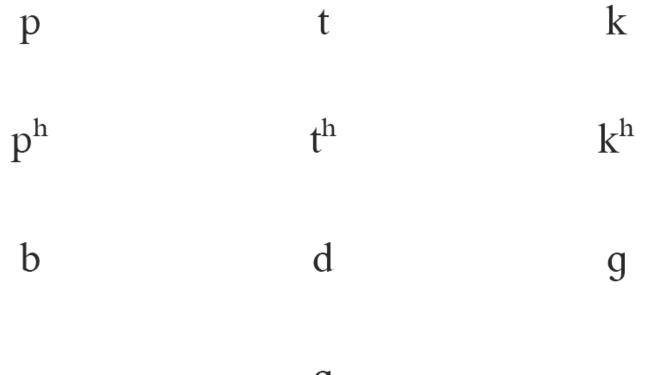Artificial language 8АИАЖА [$p^ha.t^ha.k^ha$]

- Subjects learn the language, then learn the "Northern Dialect", with spirantized voiced stops.

- How will subjects pronounce Aspirated Stops?

Subjects and Schedule

SUBJECT	AGE	SEX	LINGUIST	LANGUAGE BACKGROUND
S1	22	F	Yes	English
S2	22	M	Yes	English
S3	22	F	Yes	English
S4	19	F	No	English

- Two weeks to learn "Standard Variety" (SV)
- Reading Task in SV
- Presented with "Northern Dialect" (ND)
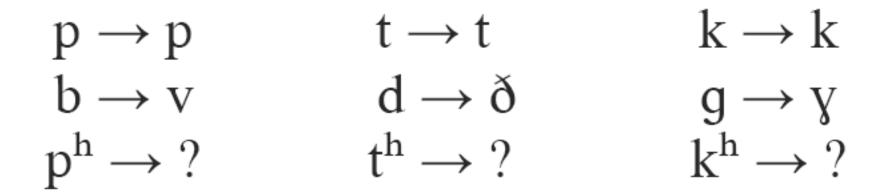- One week to learn ND
- Reading Task in ND

Inventory

p	t	k
p^h	t^h	k^h
b	d	g

ɑ

Orthography

А	/ɑ/		Ю	/d/
ꓑ	/p/		И	/tʰ/
З	/b/		9	/k/
8	/pʰ/		ſ	/g/
Л	/t/		Ж	/kʰ/

Northern Dialect

$$p \rightarrow p \qquad t \rightarrow t \qquad k \rightarrow k$$
$$b \rightarrow v \qquad d \rightarrow ð \qquad g \rightarrow ɣ$$
$$p^h \rightarrow ? \qquad t^h \rightarrow ? \qquad k^h \rightarrow ?$$
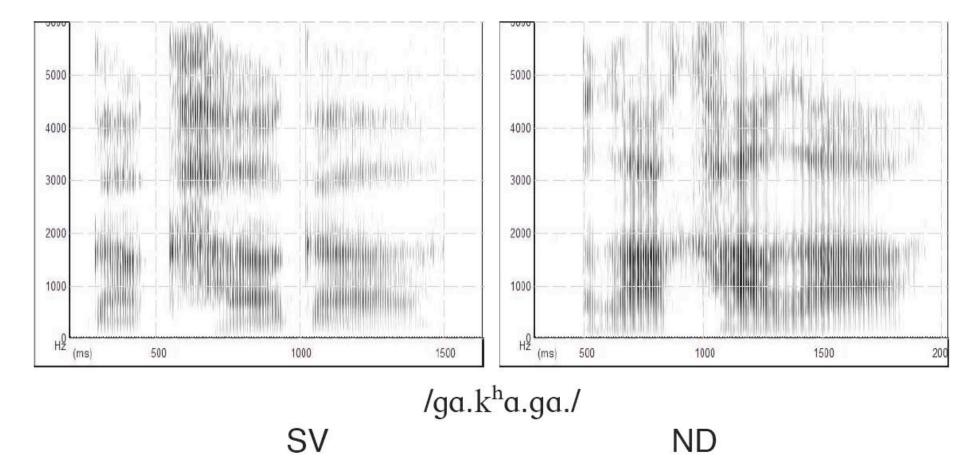
Sample

/gada ga daba gak^{h}aga/

"the fisherman shakes hands with the police officer in the house"

Standard Variety ← → Northern Dialect
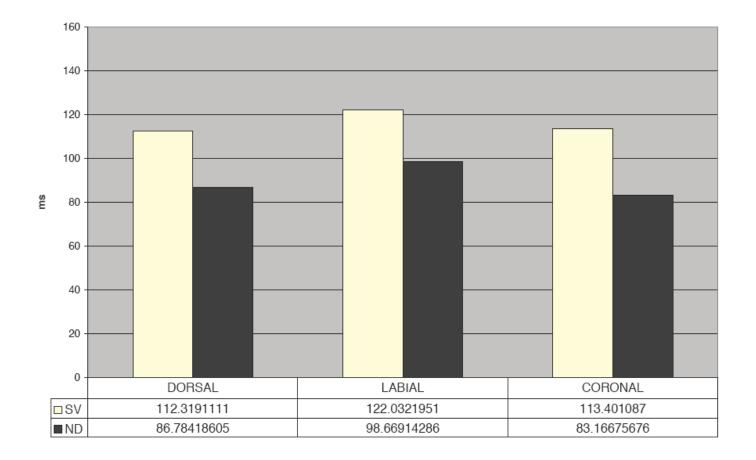
The ND aspirates lenite!

/ga.kʰa.ga./

SV   ND

Closure duration reduces

	DORSAL	LABIAL	CORONAL
□ SV	112.3191111	122.0321951	113.401087
■ ND	86.78418605	98.66914286	83.16675676

Subjects comment on the unsuccessful attempt to resist the urge to lenite aspirates

All Subjects: Aspirates "hard to pronounce" under changed circumstances

Linguists said they made conscious effort to produce aspirates as stops

Nielsen 2006

- Imitation task extends long VOT from /p/ to /k/

- Generalization involves a never-contrastive VOT diff.

- Training stimuli withhold /k/. Hi, lo frequencies.

- Stages:
  - Warm-up: read list silently
  - Baseline: read list aloud, prompted for each item
  - Listen: all 120 list items, 2x avg; /p/: 113ms VOT
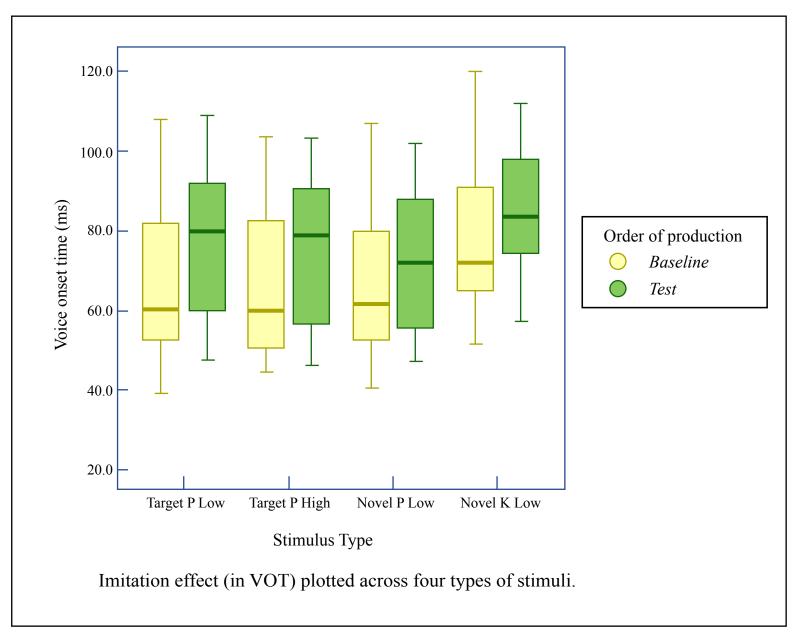 - Test: read list loud, as above

Imitation effect (in VOT) plotted across four types of stimuli.

The nature of featural categories

- Goldrick (JML 2004) reports on a set of AG results suggesting that English speakers expect /f/ and /v/ to pattern alike; also /s/ and /z/. But not /k/ and /g/.
- What makes {**f, v**}, {**s,z**} cohesive classes for English speakers?
- Shared phonetic attributes?
- Why not also {**k, g**}?
- Alternating status, suggesting equivalence?

  roof, roo[vz]; elf, el[vz]; life, alive

  pack[s], bag[z]; hou[s], hou[z], hou[z]es
- A mix of both?

Sapir

- distributional similarity as the possible basis of a phonological category

Nootka

ph	th	tsh	tSh	kh	kwh	qh
p	t	ts	tS	k	kw	q
m	n	l	j		w	
p'	t'	ts'	tS'	k'	kw'	q'
'm	'n	'l	'j		'w	

Sapir 1933, reprinted in Mandelbaum ed. 1963

Phonetic differences

- **p! is an ejective**: "synchronous closure of lips and glottal cords […] sudden release of lip closure, a moment of pause and then the release of glottal closure […] click-like character"

- **'m is a preglottalized sonorant**: "while lip closure and glottal closure are synchronous as before, the glottal closure must be released at the point of initial sonancy of the **m**."

- Spelling "difference **p!** vs. **'m** […] was not unjustified on purely phonetic grounds"

Alex Thomas

- Taught **<p!ap!i:>** but **<'ma:'mi:qsu>**
- Accepts **<p!ap!i:>**
- Volunteers **<m!a:m!i:qsu>**
- "valuable evidence for the phonologic reality of a glottalized class of consonants, which included both type **p'** (with prior release of oral closure) and type **'m** (with prior release of glottal closure)."
- basis for choosing this broader class, when a narrower one was suggested by the spelling?

Sapir: distributional parallelism

- Neither T' nor 'R can occur syllable finally.
- Suffixes that turn T into T' and R into 'R:
- **wi:nap** 'stay'
- **wi:nap'-a/a** 'stay on the rocks'
- **tlum** 'to be hot'
- **tlu'm-a/a** 'be hot on the rocks'

Language specific sound classes

- "Morphology… supports the phonologic proportion **p:p'= m:'m**[…] In other languages, with different phonologic and morpholgic understandings, such a parallel of orthography might not be justified at all and the phonetic differences that actually obtain between **'m** and **p'** would have a significantly different psychologic weighting"

Yokuts

- Both T' and 'R can occur syllable finally.
- So can all other C's: no natural class here
- Suffixes turn R into 'R, but not T into T'
- '-feature seeks R, ignoring intervening T

(8) *Glottalisation of sonorants over another consonant*

ʔilk-	ʔel'kaahin	'sing'
dull-	dol'laahin	'climb'
yawl-	yaẘlaahin	'follow'

From Howe and Pulleyblank 2001, *Phonology*

- Suppose T' and 'R share a phonetic feature: [+constricted glottis]
- Evidence was consistent with 2 classes:

  [-son, +c.g] and [+son, +c.g.]
- Spelling encouraged Alex to focus on 2 classes: <p!> vs. <'m>
- [spelling effect on analysis in <melon>, <cello>]
- Alex disregards spelling, glottal timing difference, probable auditory difference between T' and 'R, to focus just on a shared articulatory property.

Why

- because features refer primarily to constrictions in the vocal tract - not timing or auditory properties? (cf. Browman and Goldstein)

- or because Alex is enforcing feature economy? (cf. Clements: reducing the feature/segment ratio)

- or underspecification? (cf. Clements, Lahiri)

- or because sounds are categorized primarily on the evidence of *distributional* similarities?

  (Sapir, and now Mielke and others; possibly Goldrick's data).

- This is an anecdote, not a controlled experiment