A Constraint-Based Model of Gradient Phonotactics

Colin Wilson, UCLA

Department of Cognitive Science
Johns Hopkins University
February 25, 2008

Acknowledgments

- Bruce Hayes (UCLA) — Colleague, mentor, and collaborator on Hayes & Wilson (to appear)
- Donca Steriade (MIT) — Co-teacher of 'Analytic Bias in Phonology' at the 2007 LSA Summer Institute
- Adam Albright (MIT), Lisa Davidson (NYU), Shabnam Shademan (UCLA) — For data and helpful discussion
- Victoria Thatte (UCLA), Lauren Varner — Collaborators on related infant phonological learning studies

Three problems of generative linguistics

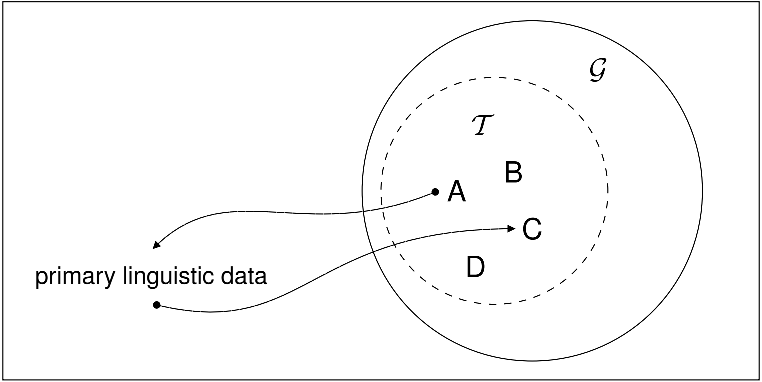
1. Descriptive adequacy: What grammars fully characterize native speakers' knowledge of language?
2. Explanatory adequacy: How are such grammars systematically acquired from positive linguistic evidence?

Chomsky (1965), Chomsky & Halle (1965)

3. Typological adequacy: Why are there strong similarities among languages: universals and near universals?

Chomsky (1986), Prince & Smolensky (1993/2004), Blevins (2004, *et seq.*)

Descriptive, explanatory, typological adequacy



- How does a native speaker learn a descriptively adequate grammar C given positive evidence drawn from an ambient grammar A that lies within the set $\mathcal{G}$ of possible grammars?
- What restricts the set $\mathcal{T}$ of typologically attested/'attainable' grammars?

Descriptive adequacy transcends observation

- An observationally adequate grammar accounts for, or 'accepts', all of the primary linguistic data (Chomsky 1965). Ex. *preen* [pɹin] is a known word of American English, and so must be accepted by any native speaker's grammar
- A descriptively adequate grammar must be observationally adequate and correctly predict the grammatical status of novel utterances (Chomsky 1965, Chomsky & Halle 1965). Ex. *pleen* is a possible word of AE, unlike **pneen* or **rpin*

The first goal of this talk is to provide a framework within which descriptively adequate phonotactic grammars can be written. Data bearing on descriptive adequacy comes from native speaker intuitions and a variety of experimental results.

Explanatory adequacy is about selection

- The problem of explanatory adequacy is one of selecting descriptively adequate grammars.
- Ex. (slightly adapted from Chomsky & Halle 1965)

	preen	*pleen*	*pneen*	*pteen*
G1: A consonant in the environment #[p _ in] must be the retroflex liquid ɹ		*	*	*
G2: A consonant in #[p _ in] must be a liquid, { ɹ, l }			*	*
G3: A consonant in #[p _ in] must be a sonorant, {ɹ, l, m, n, ŋ, w, j, . . . }				*

The second goal of the talk is to provide a function-level theory (Marr 1982) of how descriptively adequate phonotactic grammars of the type proposed are acquired. Application of the theory to child and adult learning is a topic of current research.

Overview of the talk

§1 Phonotactic knowledge is cross-classifying[1] and gradient[2].

[1] Clements & Keyser 1983, Pierrehumbert 2003. [2] Greenberg & Jenkins 1964, Scholes 1966, Pertz & Bever 1975, Ohala & Ohala 1986, Coleman & Pierrehumbert 1997, Hammond 1999, 2004, Frisch et al. 2004, Treiman et al. 2000, Bailey & Hahn 2001, Boersma & Hayes 2001, Moreton 2002, Coetzee 2004, Buchwald 2005, Albright 2006, Davidson 2006, McClelland & Van der Wyck 2006, Berent et al. 2007

§2 Constraints weighted according to the principle of maximum entropy[3], and selected by comparing observed and expected ratios[4], provide descriptively adequate grammars that capture cross-classification and gradience.

[3] Della Pietra et al. 1997, Jelinek 1999, Berger 'All That' ms., Eisner 2001, 2002, Goldwater & Johnson 2003, Jäger 2004, Keller 2005; see also Smolensky 1986. [4] Pierrehumbert 1994, Frisch & Zawaydeh 2001

Outline

Phonotactic knowledge

- Every language imposes restrictions on the sounds and sound combinations that make up well-formed morphemes and words. These are the 'phonotactics' of the language.

  < Latin *tactica* < Greek *taktika* 'matters pertaining to arrangement' , also Greek *taktikē (tekhnē)* '(art) of deploying forces in war', < PIE **tag-* 'to set aright' (Answers.com, Online Etymology Dictionary)

- Types of phonotactic restrictions (clearly not exhaustive):
  - Possible word-initial and word-final sounds
  - Allowed consonant sequences, vowel sequences
  - Well-formed syllables and other prosodic constituents
  - Stress and pitch accent

Native speakers internalize the restrictions in the form of a phonotactic grammar as part of their knowledge of phonology.

Why study phonotactics?

Reference to phonotactics pervades work in theoretical phonology, psycholinguistics, acquisition, and related areas:

- Phonotactics motivate phonological processes[1] and may reveal subsyllabic constituency[2].

  [1] Silverstein 1974, Kenstowicz & Kisseberth 1977, 1979:ch.10, Goldsmith 1993, Steriade 1982, Pater & Tessier 2003, Tesar & Prince 2003, Bakovic 2005; [2] Kessler & Treiman 1997, Lee & Goldrick 2007

- Phonotactics influence speech perception[3] (‘illusions’, ‘deafness’) and segmentation of continuous speech[4].

  [3] Dupoux et al. 1997, 1998, 1999, Hallé et al. 1998, Pitt 1998, Moreton & Amano 1999, Moreton 2002, Hay et al. 2004, Kabak & Idsardi 2004. [4] Norris et al. 1997, Suomi et al. 1997, McQueen 1998, McQueen et al. 2001, Norris 2002

- Phonotactic well-formedness affects adult recognition[5] and production[6] of words and nonwords.

  [5] Vitevitch & Luce 1999, Vitevitch et al. 1997, 1999, 2004. [6] Vitevitch et al. 2004, Vitevitch & Luce 2005, Davidson 2006, Berent et al. 2007, Goldrick & Larsen (in press)

Why study phonotactics? (continued)

Reference to phonotactics pervades work in theoretical phonology, psycholinguistics, acquisition, and related areas:

- Phonotactics are acquired early in development (some at 9mo. or earlier)[7], phonotactics influence infant speech segmentation[8], and phonotactically well-formed words are learned and produced more accurately by children[9].

  [7]Jusczyk et al. 1993, Jusczyk & Charles-Luce 1994, Zamuner 2006. [8]Friederici & Wessels 1993, Mattys et al. 1999, Mattys & Juszczyk 2001; see also Brent & Cartwright 1995. [9]Storkel & Rogers, Storkel 2001, 2003, 2004, Munson 2001

- Novel phonotactics can be learned rapidly by infants[10] and adults[11], providing evidence for specific learning biases.

  [10]Saffran et al. 1996, Chambers et al. 2003, Saffran & Thiesson 2003, Newport & Aslin 2004. [11]Onishi et al. 2002, Pycha et al. 2003, Wilson 2003, Moreton 2006, to appear, Finley & Badecker 2007, 2008

Why study phonotactics? (continued)

- Phonotactics vary across lexical categories[12], suggesting that they could influence syntactic acquisition and parsing.

[12] Kelly & Martin 1994, Onnis & Christiansen ms., Guion et al. 2003

- Phonotactic restrictions can persist diachronically despite large-scale lexical replacement, perhaps reflecting selective use or borrowing[13].

[13] Ferguson & Farwell 1975, Ingram 1978, Locke 1983, Martin 2007

Ex. Martin (2007) shows that the dispreference for identical liquids, *[l . . . l] and *[ɹ . . . ɹ], has persisted in English for at least 1500 years, finding evidence for it in: neologisms, popular names (e.g., *Gerard*, *Leila* are infrequent across the decades), drug brand names, Fantasy RPG names!

Why study phonotactics? (because it's fun)

- Avoid difficult words, and know why you do it.
  Avoidance study of adolescents and adults (Locke 1982)
  Q: What words do you avoid because they are awkward or hard to say?
  A: *seminary, animal, minimum; burglar, auxiliary, behavioral; hypothesis, surreptitious, thesaurus*
- Stay one step ahead of the paparazzi.
  'Anyone interested in syllable contact and metathesis should take note of the current news reports about Britney Spears. Her manager, whose real name is apparently Sam Lutfi, is frequently called Sam Lufti by reporters. Currently, "Sam Lutfi" gets 76,700 Google hits, while "Sam Lufti" gets 60,000.'
  — Post by phonologist Nancy Hall, 02/01/2008 on phonoloblog (http://camba.ucsd.edu/phonoloblog/)

Criteria for an integrated model of phonotactics

A model of phonotactics that is responsible to linguistic, psycholinguistic, and acquisition data will:

- be *expressive* enough to provide descriptions of the phonotactic patterns found in natural languages, accounting for key properties such as cross-classification.
- make fine-grained, *gradient* distinctions of the kind evidenced in various types of tasks and natural behavior.
- include an explanatorily adequate theory of phonotactic *learning* (and ultimately of phonological development).

Expressiveness issues

- Two popular types of model are not sufficiently expressive to account for common phonotactic patterns, failing in particular on overlapping and long-distance dependencies.
  - Sequential models
    Ex. transitional probability grammars (Vitevitch & Luce 1999)
  - Hierarchical models
    Ex. word-/syllable-structure grammars (Coleman & Pierrehumbert 1997)
- The proposed maximum entropy model solves this problem by combining sequential relations, hierarchical structure, and other dependencies in a single, trainable grammar.

Against strictly sequential models

A strictly sequential model conditions the well-formedness of a given sound on the contiguous sequence of preceding symbols (up to some maximum length *n*): $\boxed{x_{-n}} \ldots \boxed{x_{-2}}\,\boxed{x_{-1}} \rightarrow \boxed{x_0}$

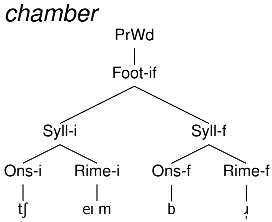
Problems

- Some attested phonotactics have **no principled upper bound** on the distance between interacting elements.

consonant harmony/disharmony (MacEachern 1999, Hansson 2001, 2004, Rose & Walker 2004)	*s . . . ʃ *tʰ . . . [h,t']	*[n,m] . . . l *t . . . d
weight-/sonority- sensitive stress (Prince 1983, Hayes 1995, Walker 1996, Bakovic 1998)	$^*\breve{\sigma}_{heavy} \ldots \acute{\sigma}_{light}$	

- Even when the interacting sounds must be ‘close’, some **intervening material can be irrelevant**. Ex. AE $^*C_iXC_i$.

Against strictly hierarchical models



A strictly hierarchical model conditions the well-formedness of a given sound on the (typically prosodic) structure above it.

Problems

- Many phonotactics operate **across hierarchically-defined constituents** (possibly as well as within them).

AE nasal place assimilation and *C_iXC_i (Pierrehumbert 1994)	*[tʃeɪn.bɹ̩], *[tʃeɪŋ.bɹ̩] *[l.fl, l.pl, t.st, t.str, nt.n, n.sn, . . .]
unbounded phonotactics	*see preceding slide*

- Does not account for similar patterns in distinct constituents: Ons-initial and Ons-final in Coleman & Pierrehumbert (1997).

Against holistic segments

In addition to the problems noted above, typical instances of the sequential and hierarchical models also treat sounds as holistic entities, unrelated by phonological feature specifications.
Ex. [p] and [t] are no more closely related than [p] and [ŋ] <ng>.

Problems

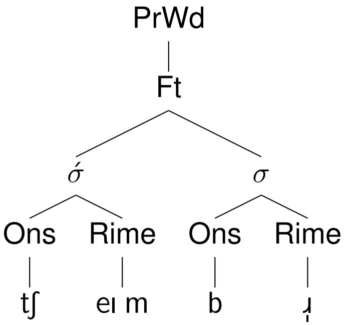
- Similar patterning of featurally-related sounds is a central discovery of phonological theory. (e.g., Jakobson & Halle 1956, Goldsmith 1979)
- Several instances of feature-based phonotactics are featured in the onset grammar presented below.

(see also esp. Frisch et al. 2004, Rose & Walker 2004, McClelland & Van der Wyck 2006)

Cross-classification in phonotactics

The typological data reviewed above motivates a theory in which **sounds have multiple overlapping, cross-classifying descriptions** for the purposes of phonotactic well-formedness.

chamber



Ex. The [m] in this word is simultaneously:

- A constituent of the *first* syllable's *rime*, as is [eɪ].
- A constituent of the *stressed* syllable, as are [tʃ] and [eɪ].
- *Adjacent* to the following [b].
- *Identical* to the [b] in place and voice.
- *Distinct* from the [b] in nasality.

These and other aspects of the sound could jointly determine its phonotactic status in the proposed constraint-based maxent model.

Cross-classification in English consonants

English has approximately 25 consonants, displayed here in modified IPA format by place of articulation, manner, and voicing

	Labial	Coronal	Dorsal	Glottal
oral stop	p b	t d	k g	(ʔ)
fricative	f v	<th>θ ð s z <sh>ʃ <s>ʒ		h
nasal stop	m	n	<ng>ŋ	
approximant	w	<r>ɹ j		
lateral		l		
affricate		<ch>tʃ <j>dʒ		

Examples of similar phonotactic behavior of related consonants:

- Only the nasal sounds [m n ŋ] are robustly found before the voiced stops [b d g], and then only under identity w.r.t. place.
- The sonorant sounds (nasals and approximants) are never followed by another sound at the beginning of the word.

Summary: Cross-classification in phonotactics

Pierrehumbert (2003): 'A combined autosegmental/metrical formalism (as is developed in Pierrehumbert & Beckman 1988) permits all of [the phonotactic patterns, CW] to be understood simply as fragments of phonological description.

- 'There is no single privileged level of analysis, and **the fragments crosscut each other in the sense that they do not stand in any fixed hierarchical relationship**.
- 'Taking the syllable as a kind of mental reference point, note that the list . . . includes patterns that are bigger or smaller than a syllable . . . ; **syllables that happen to be diphones**; **syllable junctures**, containing just the end of one syllable and the beginning of the next; and **consonantal projections** that abstract across variations in syllable structure' (p. 192, bullets and emphasis added).

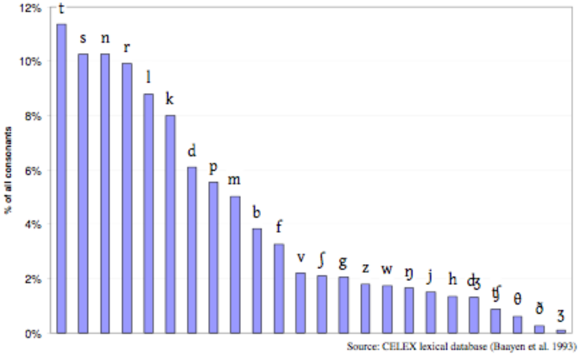
Gradience issues

- Traditional generative models of phonotactics impose a binary distinction between legal and illegal structures.
  Ex. [bl, bɹ] *[bw, bn, bz, bd, lb]
- But gradience is found in every study that allows speakers to rate forms on a scale, or averages binary responses, or reports percent correct in a production or other task.
  Ex. [bl, bɹ] > [bw] > [bn] > [bd, bz] > [lb]

  example combines Berent et al. (2006) and Albright (2007); see also Greenberg & Jenkins 1964, Scholes 1966, Pertz & Bever 1975, Ohala & Ohala 1986, Coleman & Pierrehumbert 1997, Treiman et al. 2000, Bailey & Hahn 2001, Buchwald 2005, Davidson 2006, McClelland & Van der Wyck 2006, Shademan 2007
- The failure of traditional models to consider gradience may have fostered the view that only lexical similarity, or simple sequential/hierarchical models, could possibly be ‘psychologically real’.

Gradience in lexical distribution

Even putting aside the experimental studies, it is striking that generative work has largely ignored *non-categorical* lexical distributions. Ex. AE consonants (from Martin 2007):



see also Frisch et al. 2000, Frisch & Zawaydeh 2001, McClelland & Van der Wyck 2006, Coetzee & Pater 2008

Gradience in grammar and performance

- There is no inherent contradiction between traditional (or more recent) grammar formalisms and gradient intuitions or distributions, and many connections are being forged.

  Boersma & Hayes 2001, Davidson 2003, Albright 2007, 2008, Coetzee & Pater 2008, Dmitrieva et al. 2008

- Linking gradient grammar to variable performance is largely exploratory at present (but see Davidson 2003, 2006, Buchwald 2005). Two prejudices of your speaker:
  - Accept all systematic data as potentially relevant: even from metalinguistic, pointless tasks such as wordlikeness rating.
  - Adopt simple linking hypotheses, such as rating / production $\propto$ probability, as proxies for full performance models.

Outline

Constraints and weights

A maximum entropy ('maxent') grammar is composed of

- a set of constraints $\{C_1, C_2, \ldots, C_n\}$,
- each of which has a real-valued weight $\{w_1, w_2, \ldots, w_n\}$.

  Berger et al. 1996, Della Pietra et al. 1997, Jelinek 1999; tutorial: Klein & Manning 2003

- All of the constraints considered here are negative (prohibitions), therefore the weights will be restricted to non-negative values: larger weight $\Rightarrow$ stronger prohibition.

Ex.

	*C] (2)	*V] (1)
[ta]		
[a]		*
[tak]	*	
[ak]	*	*

Evaluation of forms

A phonological representation x is evaluated by

- summing its weighted constraint violations

  $$h(x) = \sum_{i=1}^{n} w_i \cdot C_i(x)$$

- negating the result and raising e (≈ 2.718) to that power

  $$\Phi(x) = e^{-h(x)} = \exp[-h(x)]$$

  (negation ensures that worse violators have smaller Φs)

Ex.

x	*C] (2)	*V] (1)	$h(x)$	$\Phi(x)$
[ta]			0	$\exp[0] = 1$
[a]		*	1	$\exp[-1] \approx .368$
[tak]	*		2	$\exp[-2] \approx .135$
[ak]	*	*	3	$\exp[-3] \approx .050$

Gradient phonotactic well-formedness

The well-formedness of representation x according to a maxent phonotactic grammar of weighted constraints is directly proportional to $\Phi(x)$.

- Technically, well-formedness is equated with probability
  $$\Pr(x) = \Phi(x)/Z,$$
  where Z is a normalizing constant.
- In comparing predictions with responses, we ignore the constant factor Z (this does not affect correlations).
- In fitting human results we also often introduce a single parameter T and write
  $$\text{well-formedness}(x) \propto \Phi(x)^{1/T} = \exp[-h(x)/T]$$

Why maximum entropy?

Maximum-entropy grammars have a number of desirable properties, especially for phonotactics:

- gradience follows from the basic probabilistic structure.
- constraints can be arbitrary functions from representations to violations, allowing overlapping and cross-classification.
- weights are guaranteed to be optimal (rational), in the sense that they maximize the entropy of the system / maximize the probability of the data, given the constraints.
- strongly connected to Harmony Theory (Smolensky 1986) (connections to stochastic HG/OT still open AFAIK).

Why 'maximum entropy'

Maximum-entropy grammars give the flattest, most uniform probability / well-formedness distributions that are compatible with the constraints and observations.

see derivations in Cover & Thomas 1991, Smolensky 1986, Della Pietra et al. 1997, Jelinek 1999

Example: [Ca] forms over a fixed consonant inventory

- G0 = { }
  All [Ca] forms ([pa, ta, ka, ma, na, ŋa, ...]) have the same probability / well-formedness, determined by $\exp[0] = 1$.
- G2 = { *[ŋ (1)}
  All [Ca] forms except ŋa have the same probability / well-formedness: the probability taken from ŋa is equally divided among the other forms.

Hayes-Wilson constraint learner

- Given: A fixed set of segments with feature classifications

  segment categories and features could perhaps be learned as well; see Lin 2002, Mielke 2003

- Return: A grammar of negative, weighted constraints
  - Constraints have a simple form: essentially *X, *XY, *XYZ (where X, Y, Z are natural classes defined by the features).
  - Constraints are added one at a time to an initially empty grammar; weights are readjusted at each step.
  - Constraints are selected to penalize sounds and sequences that are **observed in the data more rarely than expected**.

Preview: Phonotactics of English consonants

English has approximately 25 consonants, displayed here in modified IPA format by place of articulation, manner, and voicing

	Labial	Coronal	Dorsal	Glottal
oral stop	p b	t d	k g	(ʔ)
fricative	f v	<th>θ ð s z <sh>ʃ <s>ʒ		h
nasal stop	m	n	<ng>ŋ	
approximant	w	<r>ɹ j		
lateral		l		
affricate		<ch>tʃ <j>dʒ		

- Free combination of two at the beginning of the word would give $25^2 = 625$ possibilities, but only about 30 (5%) occur.
- Free combination of three gives $25^3 = 15,625$ possibilities, but only 5 or 6 occur initially: spɹ, stɹ, skɹ, spl, skw, (skl).
- Pierrehumbert (1994) finds only 50 CCC sequences morpheme-medially (less than 1% of expected 8708).

Consonant features

	wb	cons	son	cont	nasal	liq	voice	lab	cor	ant	strid	lat	dor	high	back
p	-	+	-	-			-	+							
b	-	+	-	-			+	+							
t	-	+	-	-			-		+	+	-				
d	-	+	-	-			+		+	+	-				
k	-	+	-	-			-						+		
g	-	+	-	-			+						+		
f	-	+	-	+			-	+							
v	-	+	-	+			+	+							
θ	-	+	-	+			-		+	+	-				
ð	-	+	-	+			+		+	+	-				
s	-	+	-	+			-		+	+	+				
z	-	+	-	+			+		+	+	+				
ʃ	-	+	-	+			-		+	-	+				
ʒ	-	+	-	+			+		+	-	+				

Consonant features

	wb	cons	son	cont	nasal	liq	voice	lab	cor	ant	strid	lat	dor	high	back
tʃ	-	+	-	-			-		+	-	+				
dʒ	-	+	-	-			+		+	-	+				
m	-	+	+		+	-		+							
n	-	+	+		+	-			+	+	-				
ŋ	-	+	+		+	-							+		
ɹ	-	-	+			+			+	-	-				
l	-	+	+			+			+	+	-	+			
w	-	-	+			-		+						+	+
j	-	-	+			-								+	-
h	-	-	-	+			-								

Two more features for the glides: [+high] j w, [+back] w, [–back] j.

Note extensive use of underspecification (aiming to limit the number of classes): e.g., [±voice] → [–son], [±liq(uid)] → [+son].

Complement natural classes

For every natural class X, we also allow the learner to consider ˆX, the complement of X.

- Complement class ˆX contains all and only the segments not in X (i.e., ˆX = $\Sigma \backslash$X).
- Complementation allows some phonotactics to be expressed in a unitary fashion:
    - *[ˆs][-sonorant]
      'no segment except [s] can precede an obstruent'
      (e.g., *[db], *[lb], *[zb])
    - *[+lab][ˆ+liquid]
      'a labial cannot precede anything but a liquid'
      (e.g., *[bw], *[bn])

Constraint learning

Consider two classes, say [+sonorant] and C (any consonant). Should the learner add the constraint *[+son]C to the grammar?

- **Observed**
  Suppose the learner knows that **O**([+son]C) = 0 based on experience with the primary linguistic data.
- **Expected**
  The learner can estimate **E**([+son]C) by randomly generating clusters with its current grammar (initially { }).

Constraint learning (continued)

- If violations of *[+son]C are observed more rarely than expected, as measured by

$$\textit{accuracy} = \frac{\mathbf{O}([+\text{son}]\text{C}) + \epsilon}{\mathbf{E}([+\text{son}]\text{C}) + \epsilon}$$

then *[+son]C is a viable new constraint.

Note that $\mathbf{O}([+\text{son}]\text{C}) < \mathbf{E}([+\text{son}]\text{C})$ 'rarer than expected'
$\Rightarrow \frac{\mathbf{O}([+\text{son}]\text{C})+\epsilon}{\mathbf{E}([+\text{son}]\text{C})+\epsilon} < 1.0$, with *smaller* values for larger differences.

With $\epsilon = 1$: $\frac{0+\epsilon}{1000+\epsilon} = .0009$, $\frac{100+\epsilon}{1000+\epsilon} = .10$, $\frac{0+\epsilon}{10+\epsilon} = .09$, $\frac{5+\epsilon}{100+\epsilon} = .06$

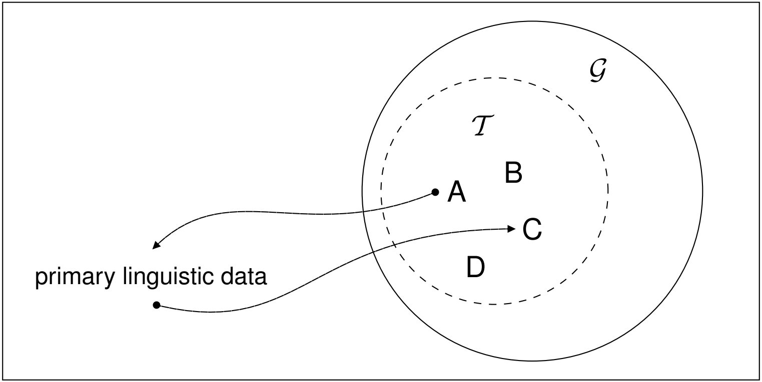
Constraint learning (continued)

Now consider *all possible* constraints over the given set of natural classes $\sim$ 1 million constraints for 98 natural classes. How does the learner select a member of this large space?

Selection heuristics prefer **accurate and general** constraints:

- From among all of the constraints with *accuracy* $< \alpha \leq 1$, where α increases over the course of learning,
- select the constraint that has the *fewest* natural classes,
- and among those select the one with the *largest* classes.

Toward descriptive and explanatory adequacy



- How does a native speaker learn a descriptively adequate grammar C given positive evidence drawn from an ambient grammar A that lies within the set $\mathcal{G}$ of possible grammars?

Case study: English word-initial onsets

- A classic topic of generative phonotactics (Clements & Keyser 1983), many hand-written analyses to compare with our learner's output.
- Growing body of experimental work on native speakers' intuitions, productions, and perceptions of novel onsets (and production of existing clusters under impairment).
- Native speaker's knowledge clearly extends beyond the set of attested onsets. Does this compel us to adopt universal constraints (e.g., constraints on sonority sequencing)?

The learning data (with type frequencies*)

k 2764, ɹ 2752, d 2526, s 2215, m 1965, p 1881, b 1544, l 1225, f 1222, h 1153, t 1146, pɹ 1046, w 780, n 716, v 615, g 537, dʒ 524, st 521, tɹ 515, kɹ 387, ʃ 379, gɹ 331, tʃ 329, bɹ 319, sp 313, fl 290, kl 285, sk 278, j 268, fɹ 254, pl 238, bl 213, sl 213, dɹ 211, kw 201, stɹ 183, θ 173, sw 153, gl 131, hw 111, sn 109, skɹ 93, z 83, sm 82, θɹ 73, skw 69, tw 55, spɹ 51, ʃr 40, spl 27, ð 19, dw 17, gw 11, θw 4, skl 1

*counts from the CMU Pronouncing Dictionary: http://www.speech.cs.cmu.edu/cgi-bin/cmudict

Some excluded clusters

- pw (*poitier, puebla, pueblo, puentes, puerto*)
- bw (*bois, bueno, buisson*)
- sf (*sferrazza, sforza, sphere, sphinx*)
- zw (*zwack, zwart, zwerdling, zwieback*)
- kn (*knesset, knutson*)
- All Cj (pj, bj, fj, vj, mj, kj, . . .) assumed to be parsed with j as a member of the rime, part of the diphthong [ju].

  Clements & Keyser 1983:42, Buchwald 2005

The biconsonantal learning data

C1/C2	w	r	l	nasals	obstruents
p	(lab-lab)	1046	238		
b	(lab-lab)	319	213		
t	55	515	(cor-cor)		
d	17	211	(cor-cor)		
k	201	387	285		
g	11	331	131		
θ	4	73	(cor-cor)		
f	(lab-lab)	254	29		

*type counts from CMU

The biconsonantal learning data

C1/C2	w	r	l	nasals	obstruents
p	(lab-lab)	1157	369		
b	(lab-lab)	486	389		
t	80	615	(cor-cor)		
d	11	295	(cor-cor)		
k	265	545	407		
g	7	428	148		
θ	4	107	(cor-cor)		
f	(lab-lab)	396	390		

*type counts from CELEX (epl), Baayen, Pipenbrock & Gulikers (1995)

The biconsonantal learning data

C1/C2	w	r	l	m	n	fricatives	p	t	k
s	153		213	82	109		313	521	278
ʃ		40							
h	111								

*type counts from CMU

The biconsonantal learning data

C1/C2	w	r	l	m	n	fricatives	p	t	k
s	210		311	135	185		531	994	602
ʃ		39							
h	276								

*type counts from CELEX (epl)

Pertz & Bever 1975

- Participants: children 9-11yr (N=40) and adolescents 16-19 (N=40)
- Materials: 24 pairs of #CCVC# forms with differing initial #CC and identical VC#
  mrawl – rmawl, *nlub – lnub*, *nrot – rnot*
  lnore – lgore, *rneek – rbeek*, *lnag – ldag*
- Task: Forced choice
  'Subjects were told that they were to choose, on a simplicity criterion ('easier, more likely, or more usual'), which one of two words has the initial sound cluster used in more languages in the world' (p. 154).

> 'If subjects are able to correctly predict the ordering of consonant clustering in the hierarchy [of cluster markedness] which are **not within their experience**, this would offer evidence of an internal basis for the universal hierarchy' (p. 150, emphasis in original).

Pertz & Bever 1975

- Results (mean number of A responses)

A – B	adolescents	children
NL – LN	2.75, $p < .001$	2.075, n.s.
LN – LD	2.55, $p < .001$	2.25, n.s.

- Pertz & Bever interpret the age-group difference in terms of linguistic development. An alternative to consider is that children's responses are simply noisier in such tasks.

Berent et al. (2006)

- Experiment 1 (N=16)
  - Materials: 90 #CCVC#, #CəCVC# items; auditory presentation
  - Task: Judge syllable count (1 vs. 2)
  - Results (approx. % correct):

    bnif (63) > *bdif* (28) > *lbif* (15)
- Experiment 3 (N=30)
  - Task: Same-different judgment
  - Results (∼ % correct): *bnif – bənif* (68) > *bdif – bədif* (31) > *lbif – ləbif* (31)

> ‘Our findings demonstrate that English speakers manifest sonority-related differences despite the lack of lexical evidence, either direct (i.e., the existence of the relevant onsets in the English lexicon) or indirect (the statistical co-occurrence of segments in English words)’ (p. 35)

Berent et al. (2006)

Controls

- Naive transcription of #CCVC# (fn. 5): medial epenthesis (37.4%) more freq. than substitution (5.18%) or prosthesis (1.11%).
- Experiments 2 and 3: native Russian speakers readily perceive #CC - #CəC differences across all of the materials.
- Experiment 6: native English speakers can perceive all #CC - #CəC differences tested when the task induces focus on epenthesis.

Albright 2007

- Materials: 30 monosyllabic nonwords with *p-/b-* initial clusters; rhymes were controlled for neighborhood density and bigram prob + 170 fillers (also nonwords, 70 legal).
- Task: Repetition to auditory presentation, followed by judgment on a scale: 1 (“Completely impossible”) — 7 (“fine”)
- Results (ratings and % correct repetition)

  bl > *br* > *bw* > *bn* > *bd, bz*

> ‘A much more interesting kind of fact is when speakers prefer one unattested sequence over another: **bnick* > ***bdick*, ***bzik*. In such cases, the preference that we observe could not be due to the fact that there are more *#bn* words than *#bd* or *#bz* words, since there are no words that begin with any of these clusters. Ultimately, what we would like to know is to what extent speakers’ preferences are learned (directly or indirectly) from the data of English, and to what extent they reflect prior, universal biases’ (p.1).

Summary of experimental findings

Gradient differences among non-existing onsets

- Pertz & Bever 1975
  NL $>$ LN $>$ LD
- Berent et al. 2006
  BN $>$ BD $>$ LB
- Albright 2007a
  BL, BR $>$ BW $>$ BN $>$ BD, BZ

Do these differences necessarily reflect universal constraints (e.g., constraints referring to sonority), as opposed to learned properties of the word-initial onset system?

Grammar induced by HW learner

	constraint	weight	constraint	weight
1	*[ˆs][-sonorant]	2.07	*[+labial][ˆ+liquid]	1.44
2	*ŋ	1.22	*[-strident][ˆɹ j w]	1.11
3	*[+sonorant]C	1.49	*[ɹ j w h][ˆw]	0.72
4	*ʒ	1.01	*[-continuant,-anterior]C	0.83
5	*[ˆs][+nasal]	1.37	*C[+strident]	0.80
6	*[+cont,+voice]C	1.33	*[ˆ+cons,-son][+cor]	0.86
7	*Cj	1.50	*[+voice,+strident]C	0.63
8	*C[+continuant]	1.37	*[+anterior,+strident][-ant]	0.68
9	*[-anterior][ˆɹ]	1.15	*[ˆs][+cons,+labial]	0.79
10	*C[+voice]	1.26	*[-cont,+coronal][ˆ-cons,+sonorant]	0.62

Remarks on the grammar

- Quite similar to Clements & Keyser's (1983) hand-crafted grammar (see Hayes & Wilson, to appear for details).
- General constraints that rule out many clusters, such as *[^s][-sonorant] and *[+sonorant]C, receive large weights.

Remarks on the constraints

- C9 *[-anterior][ˆɹ] (1.15) is violated by every ʃ-initial cluster except [ʃɹ].
- C12 *[-strident][ˆɹ j w] (1.11) allows [tɹ] and [dɹ] but is violated by *[tl] and *[dl].
- C13 *[ɹ j w h][ˆw] (0.72) allows [hw] (present in BH's dialect).
- C14 *[-continuant,-anterior]C (0.83) is violated by clusters that begin with [tʃ] and [dʒ].
- C18 *[+anterior,+strident][-anterior] (0.68) bans *[sɹ].

Analysis of Pertz & Bever 1975

- Review: NL $>$ LN $>$ LD (adolescent data)
- For each cluster x, compute the negative sum of weighted constraint violations, $h(x)$.
- $h(x) > h(x')$
  $\Rightarrow x$ is phonotactically better than x'

x	h(x)	*[ˆs][-son] 2.07	*[+son][] 1.49	*[ˆs][+nasal] 1.37	*[][+voice] 1.26	*[-strid][ˆ-cons,+son] 1.11	*[ˆ+cons,-son][+cor] .86
N L	-6.10	0	1	0	0	1	1
L N	-7.47	0	1	1	0	1	1
L D	-9.43	1	1	0	1	1	1

Also: all three violate *[ˆ+cons,-son][+cor] (.86) once. Recall that nasals are [0voice].

Analysis of Pertz & Bever 1975

Explication of the differences:

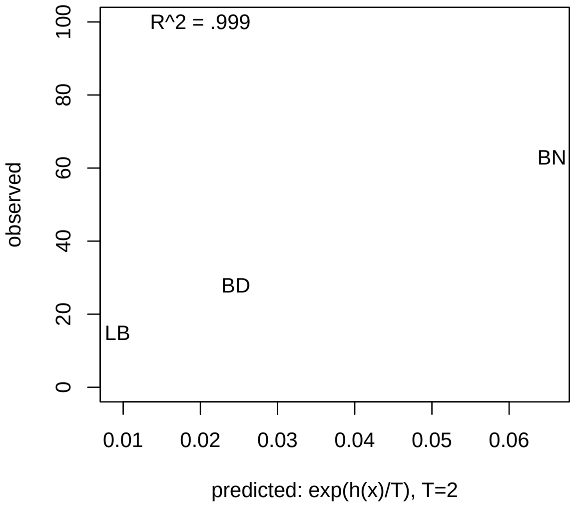
- A healthy variety of segments appear before [l] in the learning data: [p b k g f s].
  - There can be no broad and strong *Cl.
  - NL looks semi-plausible, extends a general pattern.
- Fewer segments appear before [n]: in fact only [s].
  - Succinct *[^s][+nasal] can do the job, and forcefully.
- All obstruents prefer to be at the beginning of the onset, and this preference is strongest for voiced obstruents.

Analysis of Berent et al. 2006

- Review: *bnif* (63) $>$ *bdif* (28) $>$ *lbif* (15) ($\sim$ percent correct on syllable count task)

x	hx	*[ˆs][-son] 2.07	*[+son][] 1.49	*[ˆs][+nasal] 1.37	*[][+voice] 1.26	*[+lab][ˆ+liquid] 1.44	*[-strid][ˆ-cons,+son] 1.11	*[ˆs][+cons,+LAB] .79
B N	-5.451	0	0	1	0	1	0	0
B D	-7.414	1	0	0	1	1	0	0
L B	-9.36	1	1	0	1	0	1	1

Analysis of Berent et al. 2006

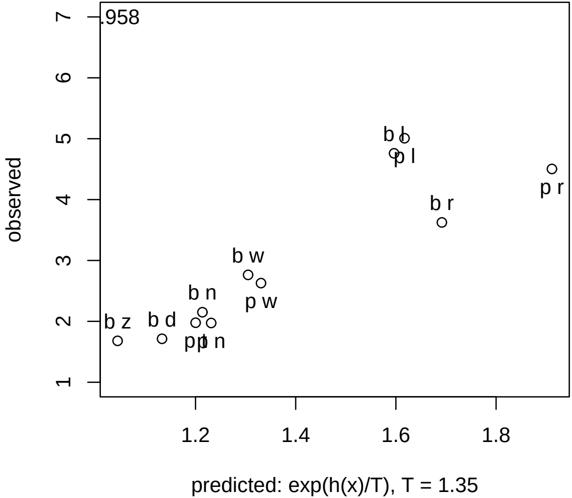


Analysis of Albright 2007

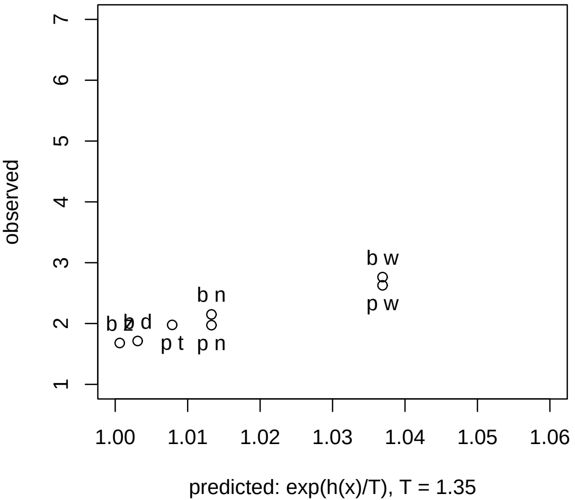
- Review (mean ratings for *b*-initial clusters): *bl* (4.76), *br* (3.62) $>$ *bw* (2.76) $>$ *bn* (2.15) $>$ *bd* (1.71), *bz* (1.68)

x	hx	*[^s][-son] 2.07	*[^s][+nasal] 1.37	*[][+cont] 1.37	*[][+voice] 1.26	*[+LAB][^+liquid] 1.44	*[][+strid] .80
B L	-2.646	0	0	0	0	0	0
B R	-2.646	0	0	0	0	0	0
B W	-4.084	0	0	0	0	1	0
B N	-5.451	0	1	0	0	1	0
B D	-7.414	1	0	0	1	1	0
B Z	-9.58	1	0	1	1	1	1

Analysis of Albright 2007a

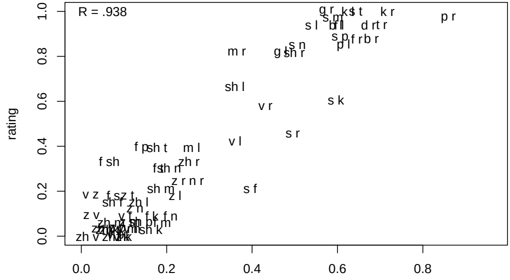


Analysis of Albright 2007a



Analysis of Scholes 1966

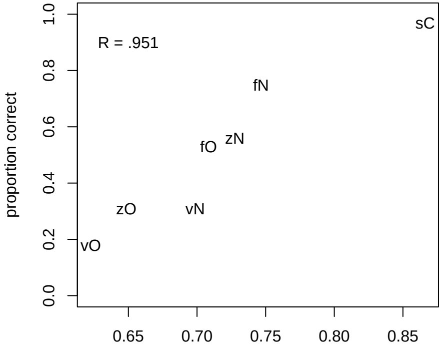
yes/no ratings of 66 monosyllabic nonwords elicited from 7th graders (N=33)



Analysis of Davidson 2006

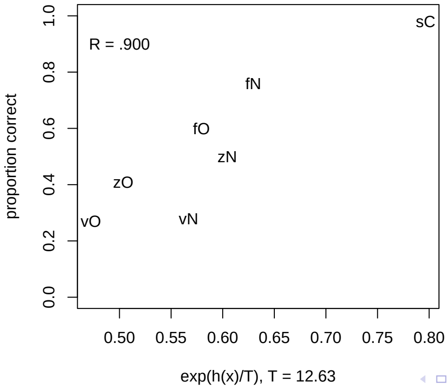
Experiment 1 (N=20): repetition of 96 Czech nonwords with native and non-native clusters.

Clusters: sm sn sf sp st sk (sC), fm fn (fN), zm zn (zN), fs fp ft fk (fO), vm vn (vN), zv zb zd zg (zO), vz vb vd vg (vO)



Analysis of Davidson 2006

Experiment 2 (N=20): repetition of 96 Slovak nonwords beginning with #CC (or #CəC).



Summary of word-initial onset analysis

The Hayes-Wilson constraint learner induces an onset grammar that captures the intuitions and performance of native speakers across a range of novel sequences.

- The grammar satisfies the requirement of descriptive adequacy to a substantial degree.
- No strong experimental evidence for including universal hierarchies of articulation or perception (such as sonority).
- Remaining difficult cases for the learner involve relative acceptability of voiced-fricative + nasal clusters.

Is there a simpler alternative?

GRM Library (Allauzen, Mohri & Roark 2003; http://www.research.att.com/ fsmtools/grm/)

- Construction of stochastic *n*-gram models that mix dependencies of several lengths
- No constraint selection stage, learning is fast
- No features or cross-cutting natural classes

Predictions (max n = 3)
bd (-12.84) $>$ lb (-13.61) $>$ bn (-13.97)
bɹ (-1.63) $>$ bl (-2.50) $\gg$ bw (-13.44) $>$ bz (-16.27)

Additional case studies

- English rimes: modeling data of McClelland & Van der Wyck 2006 with learned constraints (in progress)
- Quantity-insensitive stress systems of the world's langauges (Hayes & Wilson 2006)
- Shona vowel harmony (Hayes & Wilson 2006)
- Entire phonotactic pattern of the Australian language Wargamay (Hayes & Wilson 2006)

Try it on your data with our downloadable software!

Outline

Summary

- The problems of descriptive and explanatory adequacy are fundamental to generative linguistics. What model of learning will select the speaker's grammar given the data?
- Phonotactics is an empirical domain in which work across many areas of cognitive science can be fruitfully integrated.
- The constraint-based model proposed here satisfies basic conditions of expressiveness, and predicts the gradient results of a number of studies. We are closer to a characterization of native speaker knowledge, use, acquisition.

Improving (disproving) the learner

- Learn finer-grained distinctions among attested structures.
- Formal analysis of the 'greedy' constraint selection heuristics (e.g., is the constraint learner PAC?).
- Comparison with alternatives based on stochastically ranked constraints, SRNs, lexical similarity, ...

Reigning in the typology

- The current learner will induce many phonotactic constraints that are rare or unattested in natural language (e.g., anti-sonority-sequencing, *[–sonorant][+sonorant]).

  like infants? Saffran & Thiesson 2003, Seidl & Buckley 2005

- This has the advantage of simplicity, and of revealing the power of inductive engines to match speaker knowledge.
- But only a more articulated theory that makes specific claims about universal constraints (Prince & Smolensky 1993/2004, Burzio & Waymant, in prep.), cognitive biases (Wilson 2006, Moreton 2006, Finley & Badecker 2007, 2008, Thatte 2008), or the ravages of language use (Ohala 1986, Hayes 1999, Blevins 2004) will explain typology.

Dream applications

- Learning complete phonotactic grammars of ‘difficult’ languages, such as English, Polish, Japanese.
- Extending the model to the learning of alternations.
- Exploring common principles for learning phonotactics and orthotactics — or phonology/reading/spelling in general.

Thank you!

Major references can be found in Hayes & Wilson (to appear), available on-line, or ask me for them.