Computational linguistics for studying language in people: principles, applications and research problems

Bruce Hayes
Department of Linguistics
UCLA

Fifteenth Conference on Computational Natural Language Learning
Portland Oregon
June 23, 2011

Theme

- There are many fruitful areas of interaction for computational and descriptive/theoretical linguistics.
    - The theoretician's goal of modeling language **as internalized by people** offers new and intriguing problems for computationalists.
    - Computational linguistics can provide, and is providing, valuable tools to the descriptive/ theoretical linguist.
- Two case studies:
    - Sonority projection
    - Ranked-bigram morphology

Case I:
The sonority projection effect

Sonority in consonants

- A typical arrangement of consonants by sonority:

  glides >> *liquids* >> *nasals* >> *obstruents*

  [y,w] [l,r] [m,n,ŋ] [p,t,k,b,d,g,f,s,…]

  - ➢ Sonority has (rough) acoustic correlates.

Sonority sequencing principle

- Sievers 1881; Jespersen 1904; Hooper 1976; Steriade 1982; Selkirk 1984, etc.

Sonority preferentially rises uniformly through the syllable-initial clusters, and falls uniformly through the syllable-final cluster. Large rises (resp. falls) are better.

Examples of sonority sequencing

- A **pretty good** syllable: [pla] (sonority rises [p] to [l])
- A **mediocre** syllable: [pta] ([p] and [t] tied in sonority)
- A **really terrible** syllable: [lpa] (sonority falls)

- Languages **preferentially select good-sonority syllables** for their inventories (Greenberg 1978, Berent et al. 2007)
  - ➢ Exclude poor-sonority syllables entirely
  - ➢ Make poor-sonority syllables statistically rare

The Sonority Projection Effect

- Ask an English speaker:
  - How good a syllable is [lba]? (terrible sonority violation)
  - How does it compare with [bda]? (merely bad sonority violation)
- Idea: [lba] is much worse even though during language acquisition you've never heard either one—you "project beyond" what you've heard.

Experimental work demonstrating sonority projection

- English: Pertz and Bever (1975), Berent, Steriade, Lennertz, and Vaknin (2007), Albright (2007)
- Korean: Berent, Lennertz, Jun, Moreno, and Smolensky (2008)
- Mandarin: Ren et al. (2010)

Why is there a sonority projection effect? — theoretical speculation

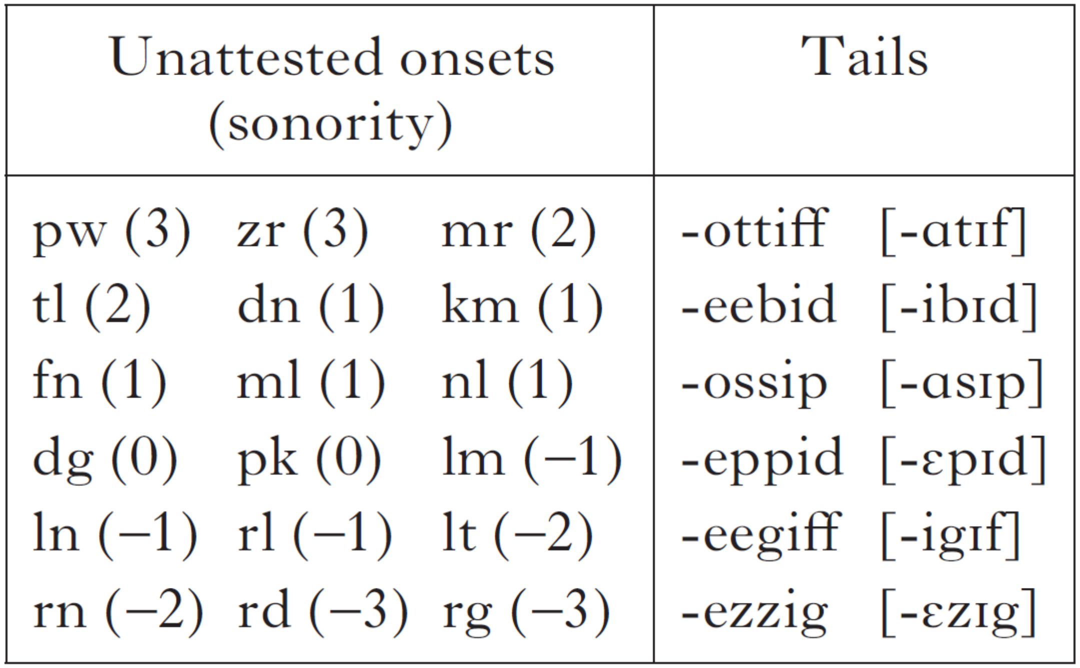
- Is it **innate**? No one has said this, but it is a logical possibility …
- Is it somehow **projected from phonetics**; i.e. avoidance of articulatory/perceptual difficulty? (cf. e.g. Hayes, Kirchner and Steriade 2004). No one has explained how this would work.
- Is it somehow **generalized from the existing clusters**? e.g. English [br, kw] etc. respect sonority, so others should.
  - ➢ Daland et al. (in press) pursue the third approach.

A computational/experimental study of sonority projection

- Reference
  - Robert Daland, Bruce Hayes, James White, Marc Garellek, Andreas Davis, and Ingrid Normann (in press) Explaining sonority projection effects. To appear in *Phonology* 28: 197–234.
- Goals
  - Do our own **ratings study**, retesting the effect.
  - Test six **computational models** of phonotactic learning to see if they could generalize sonority projection from the existing lexicon.

Experimental stimuli

- We blended nonexisting English onsets of varying sonority profile, with six "tails", e.g. *pwottiff*:

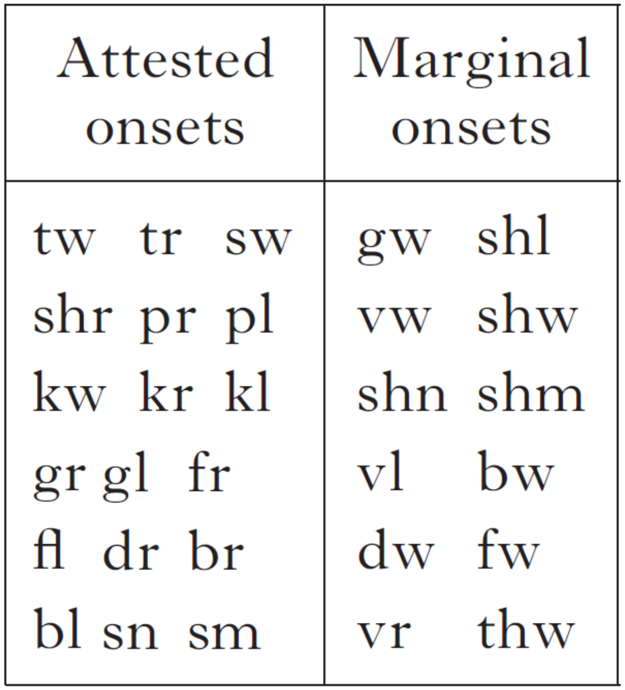


Unattested onsets (sonority)			Tails	
pw (3)	zr (3)	mr (2)	-ottiff	[-ɑtɪf]
tl (2)	dn (1)	km (1)	-eebid	[-ibɪd]
fn (1)	ml (1)	nl (1)	-ossip	[-ɑsɪp]
dg (0)	pk (0)	lm (−1)	-eppid	[-ɛpɪd]
ln (−1)	rl (−1)	lt (−2)	-eegiff	[-igɪf]
rn (−2)	rd (−3)	rg (−3)	-ezzig	[-ɛzɪg]

- Sonority "goodness scores" shown follow the sonority categories of a standard feature system (Clements 1990).

Additional stimuli

- **Attested** onsets and **marginal** (mostly loanword) onsets, attached to the same six “tails”; e.g. *twottiff*.



Attested onsets			Marginal onsets	
tw	tr	sw	gw	shl
shr	pr	pl	vw	shw
kw	kr	kl	shn	shm
gr	gl	fr	vl	bw
fl	dr	br	dw	fw
bl	sn	sm	vr	thw

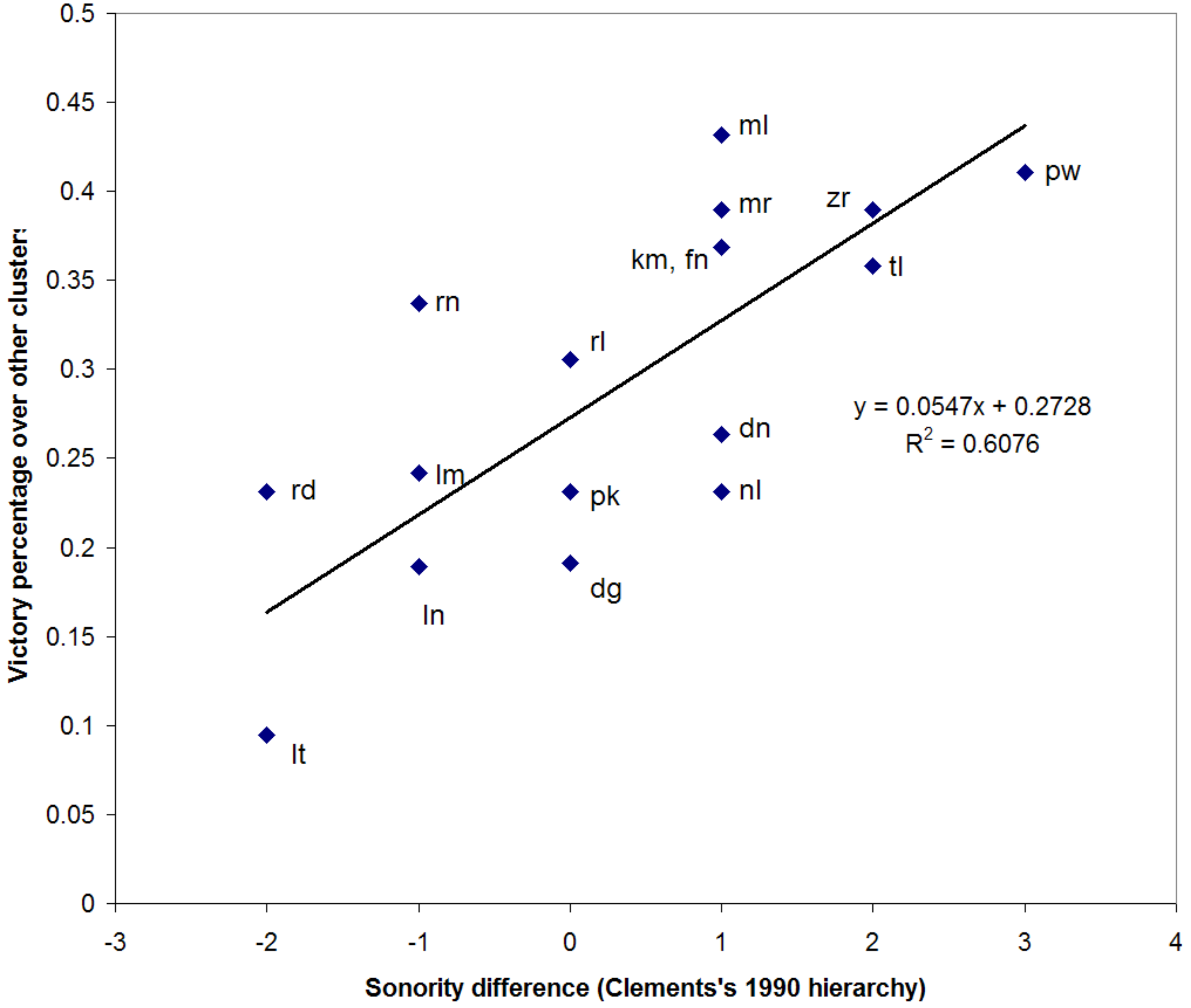
Participants and tasks

- **Participants**: from the Amazon Mechanical Turk (https://www.mturk.com)
- **2 Tasks**:
    - Rate items on a **Likert scale**, 1-6.
    - **Pairwise comparison**: all possible pairs of the 96 stimuli, i.e. which sounds "more like a typical English word"

Experiments: sample result

- Chart on next slide.
- Vertical axis: **victory percentage** for each cluster, in comparison with all other clusters
- Horizontal axis: **sonority profile** of the cluster (C2 minus C1 in the categories of Clements 1990).

Experiments showed sonority projection

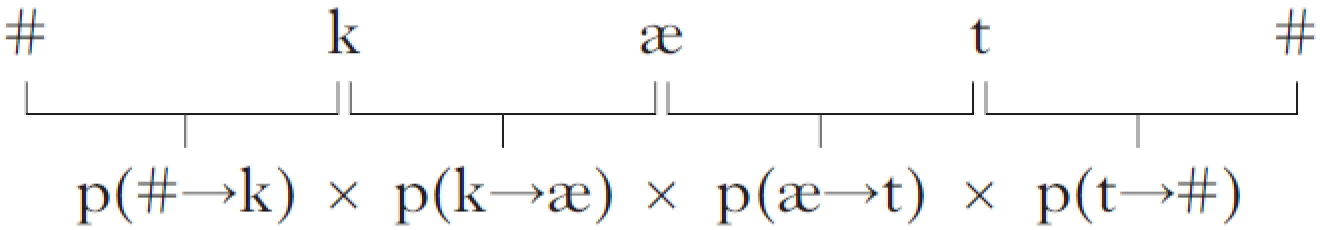


Can such intuitions be predicted from a model that learns from the lexicon?

- We tried six models; I will summarize just two.
- **Training data** for all models:
  - groomed version of the CMU corpus, words with CELEX frequencies ≥ 1, affixed and compound forms removed.

Classical bigram model

- See, e.g. Jurafsky and Martin (2000), Ch. 6.
- Calculating phonotactic probability of *cat* [kæt]:



  - ➢ Good-Turing smoothing for missing bigrams.

- Taking this as a model of human judgment: the probabilities thus derived should correlate with subject ratings.

Model with feature-based n-grams: Hayes and Wilson (2008)

- Reference:
  - (2008) Hayes, Bruce and Colin Wilson, "A maximum entropy model of phonotactics and phonotactic learning," *Linguistic Inquiry* 39: 379-440.
- This model is meant to blend ideas from traditional phonological theory and computational linguistics.

Hayes and Wilson (2008): framework

- Employs the **maximum entropy** variant (Della Pietra et al. 1997, Goldwater & Johnson 2003) of Harmonic Grammar (Legendre et al. 1990, Smolensky & Legendre 2006, Pater 2009, Potts et al. 2010).
- Probability of a form is computed from
  - its violations of a set of **constraints**
  - the **weights** of each constraint.

Formula for computing probability

- $p(\omega) = \frac{1}{Z} e^{-\Sigma_i \lambda_i \chi_i(\omega)}$, where $Z = \Sigma_j e^{-\Sigma_i \lambda_i \chi_i(\omega_j)}$

ω	a particular word
Σ_i	summation across all constraints,
λ_i	denotes the weight of the *i*th constraint,
$\chi_i(\omega)$	the number of times ω violates the *i*th constraint
Σ_j	summation across all possible words

- Z is computed with a finite state machine.

Constraint format

- A constraint consists of a unigram, bigram, or trigram of **natural classes**.
- These are defined by a standard phonological feature set, given to the model in advance.
- Example: the bigram

$$*\begin{bmatrix} -\text{sonorant} \\ -\text{voice} \end{bmatrix} \begin{bmatrix} -\text{sonorant} \\ +\text{voice} \end{bmatrix}$$

(= *[p t t͡ʃ k f θ s ʃ h][b d d͡ʒ g v ð z ʒ])

“*Don’t have a voiceless obstruent followed by a voiced obstruent.*”

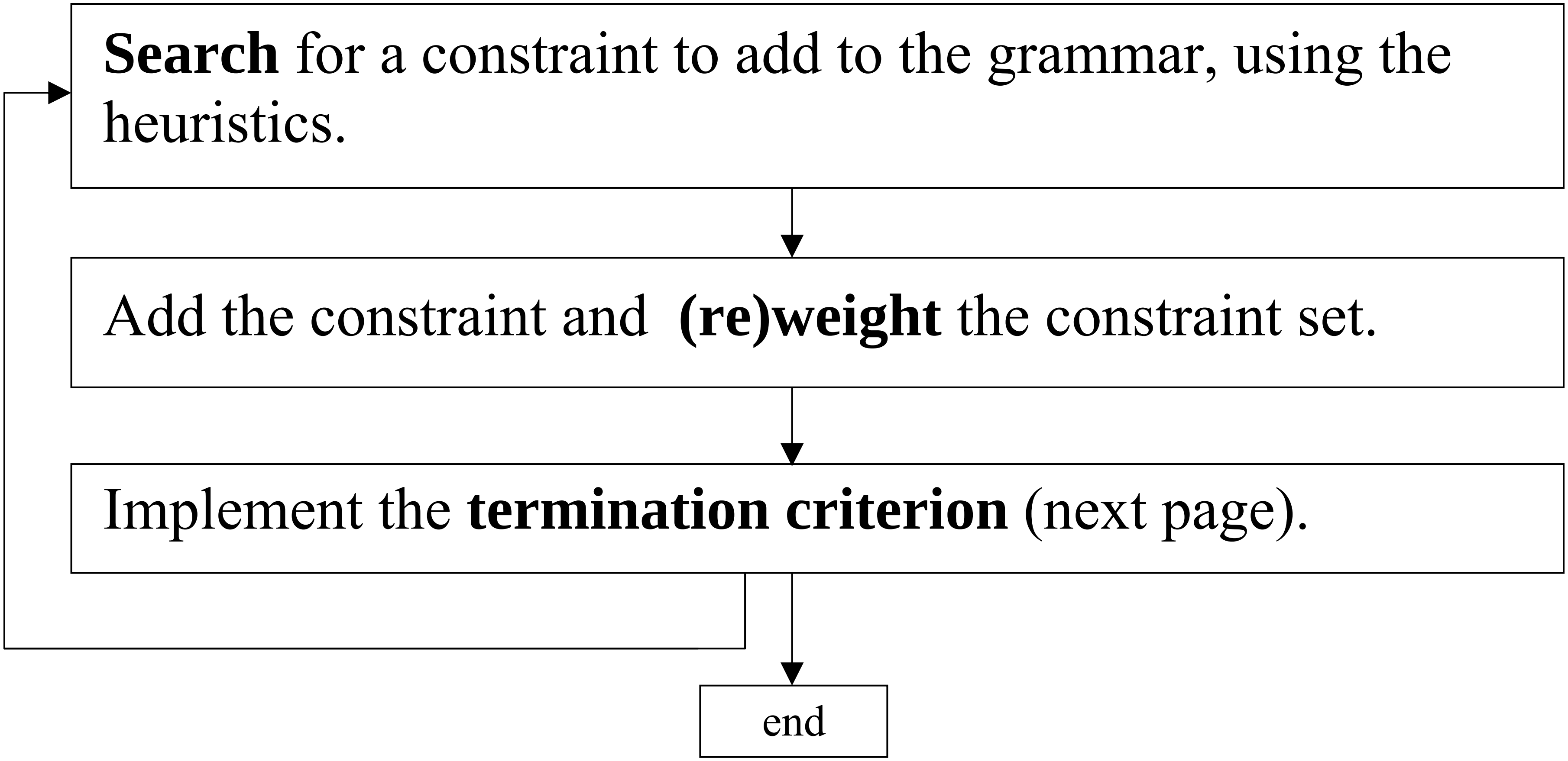
The question of search space

- The features employed define 617 distinct natural classes of sounds.
- So the number of possible constraints = $617 + 617^2 + 617^3 \approx 235$ million—small enough to work.

Picking constraints with heuristics

- Choose as follows:
  - **Fewest grams first**;
  - Among equal gram size, most **accurate** first (rising sequence of accuracy thresholds)
  - Within accuracy thresholds, most **general** first.

Overall organization of the model



Search for a constraint to add to the grammar, using the heuristics.

Add the constraint and **(re)weight** the constraint set.

Implement the **termination criterion** (next page).

end

Termination criterion used

- There are principled criteria available (e.g. upper limit for constraint accuracy) …
- … but we simply we stopped at 100 constraints
- We got similar but slightly worse results at various grammar sizes up to 350.

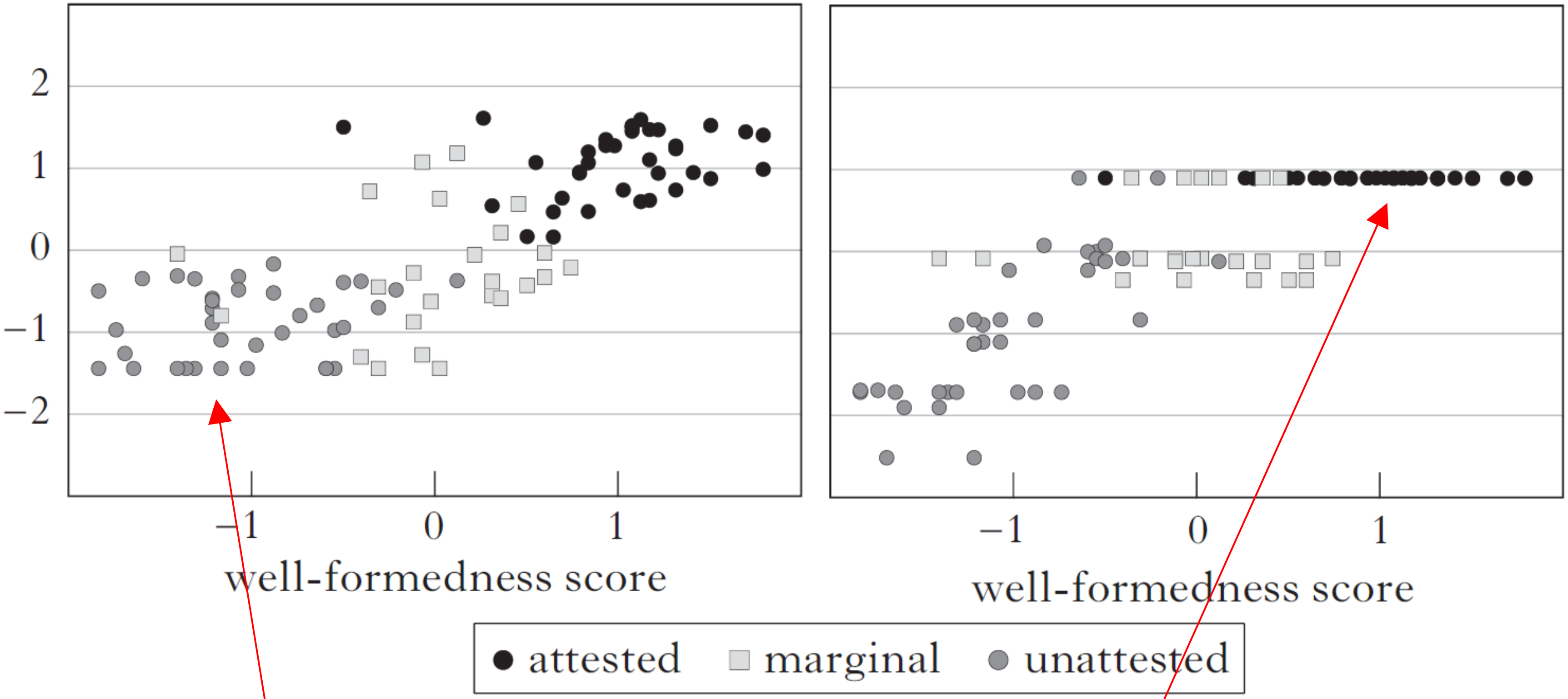
Projecting sonority: comparison of Hayes/Wilson model with classical bigrams

- We ran both models through the experimental stimuli, simulating human subjects.
- Correlation among the unattested onsets:
  - Hayes/Wilson $r = .76$ — *projects sonority*
  - Classical bigrams: $r = .22$ — *mostly doesn't*

How does the H/W model project sonority?

- As expected: it uses sonority-depicting features to generalize from the existing sonority-respecting clusters of English ([bl], [gr], [kw], etc.)
- But, an interesting wrinkle:

Scattergrams (normalized scales)



- Classical bigrams model “flattens out” for unattested onsets.
- H/W model flattens out for attested onsets.

An intuition concerning the relative strengths of these models: “figure vs. ground”

- The legal words of a language constitute a **figure** against the **ground** of all possible phoneme sequences.
- H/W model looks at the “ground” (illegal words) and penalizes large areas of it with highly general constraints.
- The bigram model looks at the figure, and makes very refined (no-features) distinctions within it — hence has little to say about the ground.

Possible lessons— model comparison

- The traditional bigram model would *never* be taken seriously by a descriptive linguist!
  - Features/natural classes are considered essential for phonological modeling.
- The failure of this model to project sonority results from its lack of features (discussion: Daland et al.)
- Yet the traditional bigram model has its virtues: covers the existing forms in great detail.
- For the future: perhaps we should try a hybrid model:
  - penalties for constraint violations
  - rewards for existing sequences

Possible lessons—benefits for descriptive linguists from computational work

- **Maxent grammars** are an extremely useful tool that descriptive linguists have borrowed from computational linguists. They offer:
    - **Total flexibility** re. content
    - **Total accuracy** in mimicking frequencies of a training set (where constraints permit)
    - Mathematical **proof of convergence**
- **Finite state machines** provide rigor and security for linguistic analysis in theories that access huge or infinite sets, as here. See Riggle (2004), Karttunen (2006), Eisner (1997, 2001, 2002)

Case II:

Weighted bigrams for morpheme ordering

Setting the scene

- Languages frequently have multiple morphemes per word (Finnish, Swahili, etc.).
- What are the principles by which these morphemes are linearly ordered?
- **Meaning** clearly plays a role, e.g. in some languages:
  - *cause to be cooked*: COOK-passive-causative
  - *be caused to cook*: COOK-causative-passive
- This is an instance of Baker's (1985) "Mirror Principle"
- However, meaning is often **overridden** by purely formal morpheme-ordering requirements.

Meaning overridden by form: Luganda (McPherson and Paster 2009)

- nyw-es-ebw-a = drink-causative-passive-final vowel

  should mean “be made to drink”

- nyw-ebw-es-a = drink- passive-causative-final vowel

  should mean “cause to be drunk”

- Only nyw-es-ebw-a is grammatical, and it has **both meanings**.
- Such **fixed orderings** are common in Bantu (Hyman 2002).

The classic account of morpheme ordering: position classes

- Wonderly (1951)'s position classes for Zoque. To make a word, pick a stem and up to one from each column.

STEM	Position Class 1	Position Class 2	Position Class 3	Position Class 4+
tah 'dig' *poy* 'run' *ken* 'look' …	-*hay* 'benefactive' -*atəh* 'indef. obj. -*ʔaŋheh* 'leave off'	-*u* 'past' -*pa* 'pres.' -*a* 'negative'	-*ək* 'where' -*məy* 'when'	(Etc. 10 classes total)

The classical account has a natural expression in Optimality Theory[1]

- For OT, see Prince and Smolensky (1993) et seq.
- **Constraints** are of the ALIGN family (McCarthy/Prince 1993)
  - ➢ ALIGN(Pos1, Left): "Assess a violation for every morpheme that precedes a Position 1 morpheme."
  - ➢ outranks ALIGN(Pos2, Left)
  - ➢ outranks ALIGN(Pos3, Left)

➢ etc.

---

[1] See Hargus and Tuttle (1997), Trommer (2003), Jaker (2006)

More on implementing the classical account in Optimality Theory

- GEN: candidates are **all possible orderings of the morphemes** in the input (*n*! for *n* morphemes).
- For the morpheme list {Stem, A, B}, the candidate set is thus

Stem-A-B, Stem-B-A, A-Stem-B
A-B-Stem, B-Stem-A, B-A-Stem

Tableau: {Stem, A, B} → [Stem-B-A]

{Stem, A, B}	ALIGN(STEM, LEFT)	ALIGN(B, LEFT)	ALIGN(A, LEFT)
☞ Stem-B-A		*	**
Stem-A-B		**!	*
B-Stem-A	*!		**
A-Stem-B	*!	**	
B-A-Stem	*!*		*
A-B-Stem	*!*	*	

- **Candidates** are sorted lexicographically by increasing violation count, respecting the ranking of the **constraints**.
- **Winner** (output of grammar) is the first in this sort.

Extension to free variation

- **Free variation** in morpheme order is surprisingly common.[2]
- Suppose *Stem-A-B* and *Stem-B-A* surface with 67/33 probability (zero for all others; e.g. **A-Stem-B*).
- We can shift to maxent grammars, assigning weights to the constraints and computing probability of candidates by the formula given earlier
  - Here Z sums across candidates, not all possible words.

---

[2] See Ryan (2010, §1)

Tableau for the free-variation case

{Stem, A, B}		ALIGN (STEM, L)	ALIGN (B, L)	ALIGN (A, L)
		10.1	**0.7**	**0**
☞ Stem-B-A	**.67**		*	**
☞ Stem-A-B	**.33**		**!	*
B-Stem-A	**0**	*!		**
A-Stem-B	**0**	*!	**	
B-A-Stem	**0**	*!*		*
A-B-Stem	**0**	*!*	*	

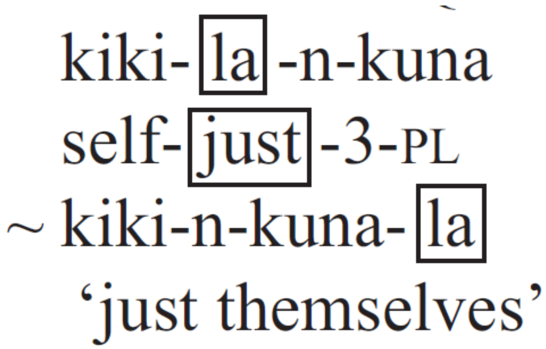
- The maxent grammar with the weights in Row 2 will derive the frequencies in Column 2.

BUT: for hard cases, ALIGN constraints work badly

- Reference:
  - Ryan, Kevin (2010) Variable affix order: grammar and learning. *Language* 86: 758-791
- Ryan points out three harder phenomena that Alignment constraints can’t cover.

Phenomenon I: Free variation moderated by "uninterruptibility"

- **X-A-B** ok, **A-B-X** ok, *__AXB__
- Real-life case: Chumbivilcas Quechua



kiki-la-n-kuna
self-just-3-PL
~ kiki-n-kuna-la
'just themselves'

- No weighting of ALIGN(X), ALIGN(A), ALIGN(B) (either direction) will work.
  - *AXB gets unwanted probability.

Phenomenon II: one morpheme “moves through a frame”

- Ryan gives a real life example from Tagalog.
- **X-A-B** ok, **A-X-B** ok, **A-B-X** ok, but nothing with B preceding A.
- No weighting of Alignment constraints works.
  - ➢ *BAX, *BXA, *XBA get unwanted probability.

Phenomenon III: Free morpheme order overridden by “gluing”

- **A-B** ok, **B-A** ok, **A-B-G** ok, ***B-A-G** bad.
  - G is “glued” to B.
- Again, Alignment fails:
  - ***BAG** gets unwanted probability.
- Example from Tagalog follows, with these morphemes:

ka-	‘telic’
RED-	‘aspect’ (realized as a copy of the following CV)
pag-	‘transitive’

Gluing example from Tagalog

- Free order:

both OK:	ma-RED-ka-tulong	ABIL-aspect-telic-help
	ma-ka-RED-tulong	ABIL-telic-aspect-help
	‘will be able to help’	

- Freedom overridden by gluing of *ka-* to *pag-*

OK:	ma-RED-ka-pag-trabaho	ABIL-asp-tel-TRANS-work
bad:	*ma-ka-RED-pag-trabaho	ABIL-tel-asp-TRANS-work
	‘will be able to work’	

- Detail: Spelling out RED. Forms would be pronounced makakatulong, makatutulong, etc.

Ryan's solution: abandon ALIGN, use Bigrams instead

- Bigrams, version I: "Assess a violation whenever a word lacks the sequence A B."
- Version II: "Assess a violation whenever morpheme A is present not followed by morpheme B."
- Version III: same as II, but "precedes" instead of "follows"
- Any of these works for Ryan; we follow him in using II.

Glueing example with bigram constraints

{A, B}		A → B	B → A	B → G
		8.0	**8.0**	**9.7**
☞ A-B	**.5**		*	*
☞ B-A	**.5**	*		*

{A, B, G}		A → B	B → A	B → G
A-B-G	**1**		*	
A-G-B	**0**	*	*	*
B-A-G	**0**	*		*
B-G-A	**0**	*	*	
G-A-B	**0**		*	*
G-B-A	**0**	*		*

(All other possible constraints are included but weighted 0.)

The previous two conundrums

- These yield to straightforward bigram solutions, too.

Language learners (mis)generalize bigrammatically: schematic example

- Early Tagalog

  ma-RED-ka- (always)
  pag-RED-pa- (always)

  perhaps because RED- started as a second-position clitic.

- Current Tagalog, long prefix string:

  ma-RED-ka-pag-pa- OR
  ma-ka-pag-RED-pa-

- A natural generalization, given the bigram constraints pag-RED and RED-pa

Language learners (mis)generalize bigrammatically: more rigorous example

- Ryan collected a large corpus of frequency data for 29 prefix combinations including RED.
- Step 1: bigrams do quite well in matching these data.
- More interesting: train on idealized data consisting of **only “first choice” forms**.
- **Train incompletely** with a gradual weight-altering algorithm.
- At the intermediate stages, *the free-variation forms of real Tagalog are generated*, with fairly accurate frequencies.

Language learners (mis)generalize bigrammatically II: the genesis of suffix copying in Bole (Chadic, Nigeria)

- Morphemes get said twice; no justification in the meaning of the form for the extra copy.
- Reference: Kevin Ryan and Russell Schuh (in progress) Suffix doubling and suffix deletion in Bole; http://www.linguistics.ucla.edu/people/hayes/205/readings/ryan_bole_handout.pdf

How Bole suffix copying works

- Required underlying configuration (suffix order shown is the expected one, based on shorter words):

  STEM + Target + Straddlee + Trigger

  - Target = suffix that gets copied
  - Straddlee = ends up flanked by copies
  - Trigger = necessary for copying to happen

- Realization:

  STEM + Target + Straddlee + Target + Trigger

An example of Bole suffix copying

ŋgòr + án + tá + án + kó
tie-plural subject-fem. sg. object-plural subject-completive
‘they tied her’

The origin of Bole suffix copying

- Related Chadic languages have the same suffixes, but no copying.
- Ryan/Schuh attribute the copying to **extension of common bigrams** (next slide).

Origin of Bole affix copying: the chain of events

- Starting point:
  - STEM-Target-Trigger (ŋgór + án + kó) was common.
  - STEM-Straddlee-Trigger (ŋgór + tá + kó) uncommon.
- Mislearning of grammar by a new generation:
  - STEM-Target and Target-Trigger highly weighted.
  - STEM-Straddlee, Straddlee-Trigger lowly weighted.
  - So STEM-Target-Straddlee-Target-Trigger becomes a plausible option.
- Basic idea is cashed out in Ryan/Schuh's partial-learning simulations.

Local summary

- Bigram theory looks like a good theory of morpheme ordering:
    - Covers cases that alignment and scope can’t cover.
    - Plausibly explains how morpheme orders evolve over time.

Research questions for ranked bigram constraint grammars I

- The **generative-capacity** question
  - Assume a symbol set S; the (infinite) set of input forms set as S*, and the set of bigram constraints defined on S.
  - What is the class of strings defined by the outputs of such grammars?
  - Does this change if we use “existence” vs. “implicational” bigrams?
  - How does this change when copying is permitted?
  - Ditto for insertion and deletion (Noyer 2001, Nunggubuyu)

Research questions for rank bigram constraint grammars II

- The **search** question:
  - Classical OT has been made formally rigorous by computational work that uses finite-state machines to insure we've considered all candidates
  - Could similar work be done for the free-ordering candidate sets needed here?
  - How does the picture change when deletion, insertion and copying are permitted?

Summing up

- Ranked-bigram constraint grammars are of interest for
    - solving previously unsolved problems in morphological analysis
    - relating to native speaker knowledge (historical change as a naturalistic wug test)
    - involving perhaps-unexplored issues of computation

Thank you

- Thanks to Kevin Ryan and Jason Riggle for helpful input.
- Author's contact information:

  bhayes@humnet.ucla.edu
  Department of Linguistics, UCLA, Los Angeles, CA,
  90095-1543

- These slides are posted at

  http://www.linguistics.ucla.edu/people/hayes/

  and include the references cited.

References

Albright, Adam (2007) Natural classes are not enough: Biased generalization in novel onset clusters. Ms., Department of Linguistics and Philosophy, MIT.

Baker, Mark (1985) The Mirror Principle and morphosyntactic explanation. *Linguistic Inquiry* 16.3: 373-415.

Berent, Iris, Donca Steriade, Tracy Lennertz, & Vered Vaknin (2007). What we know about what we have never heard: Evidence from perceptual illusions. *Cognition*, 104(3), 591-630.

Berent, Iris, Tracy Lennertz, Jongho Jun, Miguel A. Moreno, & Paul Smolensky (2008). Language universals in human brains. *Proceedings of the National Academy of Science*, 105(14), 5321-5325.

Clements, George N. (1988). The sonority cycle and syllable organization. In Dresher et al. (eds.). *Phonologica 1988*. Cambridge: Cambridge U. Press.

Della Pietra, Stephen, Vincent J. Della Pietra, & John D. Lafferty (1997). Inducing features of random fields. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 19, 380-393.

Eisner, Jason. 1997. Efficient generation in primitive Optimality Theory. In *Proceedings of the 35th Annual Meeting of the Association for Computational Linguistics*, 313–320. East Stroudsburg, Penn.: Association for Computational Linguistics.

Eisner, Jason. 2001. Expectational semirings: Flexible EM for finitestate transducers. In *Proceedings of the ESSLLI Workshop on Finite-State Methods in NLP (FSMNLP)*, ed. G. van Noord.

Eisner, Jason. 2002. Parameter estimation for probabilistic finite-state transducers. In *Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics*, 1–8. East Stroudsburg, Penn.: Association for Computational Linguistics.

Goldwater, Sharon, and Mark Johnson. 2003. Learning OT constraint rankings using a maximum entropy model. In Jennifer Spenader, Anders Eriksson, and Osten Dahl (eds.) Proceedings of the Stockholm Workshop on Variation within Optimality Theory, 111–120.

Greenberg, Joseph. H. (1978). Some generalizations concerning initial and final consonant clusters. In E. A. Moravcsik (Ed.), Universals of human language (Vol. 2, pp. 243–279). Stanford, CA: Stanford University Press.

Hayes, Bruce & Colin Wilson (2008). A maximum entropy model of phonotactics and phonotactic learning. Linguistic Inquiry, 39, 379-440.

Hayes, Bruce, Robert Kirchner, & Donca Steriade (eds.) Phonetically-based phonology. Cambridge: Cambridge University Press.

Hooper, J. B. (1976). An introduction to natural generative phonology. New York: Academic Press.

Hyman, Larry (2002) Suffix ordering in Bantu: a morphocentric approach. Yearbook of Morphology, 245-281.

Jespersen, Otto (1904). Lehrbuch der Phonetik. Leipzig and Berlin.

Jurafsky, Daniel & James H. Martin (2000). Speech Processing: An introduction to natural language processing, computational linguistics, and speech recognition (2nd edition). New Jersey: Prentice Hall.

Karttunen, Lauri (2006). The insufficiency of pencil-and-paper linguistics: the case of Finnish prosody. In *Intelligent Linguistic Architectures: Variations on themes by Ronald M. Kaplan*, Miriam Butt, Mary Dalrymple, and Tracy Holloway King (eds), pp. 287-300, CSLI Publications, Stanford, California, 2006.

Legendre, Géraldine, Yoshiro Miyata, & Paul Smolensky (1990). Harmonic grammar: A formal multi-level connectionist theory of linguistic well-formedness: an application. COGSCI 1990, 884–891.

McCarthy, John and Alan S. Prince (1993) Generalized alignment. *Yearbook of Morphology 1993*

McPherson, Laura and Mary Paster (2009) Evidence for the mirror principle and morphological templates in Luganda affix ordering. *Proceedings of the 39th annual Conference on African Linguistics*, ed. by Akinloye Ojo and Lioba Moshi, 56–66. Somerville, MA: Cascadilla.

Noyer, Rolf (2001) Clitic sequences in Nungubuyu and PF convergence. *Natural Language and Linguistic Theory* 19: 751–826.

Pater, Joe (2009). Weighted constraints in generative linguistics. *Cognitive Science*, 33, 999-1035.

Pertz, D. L. and T. G. Bever (1975). Sensitivity to phonological universals in children and adults. *Language* 51:149–162.

Potts, Christopher, Joe Pater, Karen Jesney, Rajesh Bhatt & Michael Becker (2010). Harmonic Grammar with linear programming: From linear systems to linguistic typology. *Phonology* 27, 77-117.

Prince, Alan and Paul Smolensky (1993). Optimality Theory: Constraint Interaction in Generative Grammar. Technical Report 2, Rutgers University Center for Cognitive Science.

Ren, Jie, Liqun Gao, & James L. Morgan (2010). Mandarin speakers' knowledge of the sonority sequencing principle. Presented at the 20th Colloquium on Generative Grammar at the Universitat Pompeu Fabra, Barcelona, March 18-20.

Riggle, Jason (2004) Generation, recognition, and learning in finite-state Optimality Theory. Ph.D. dissertation, UCLA.

Ryan, Kevin (2010) Variable affix order: grammar and learning. *Language* 86: 758-791.

Ryan, Kevin and Russell Schuh (in progress) Suffix doubling and suffix deletion in Bole; ms., Department of Linguistics, UCLA.

http://www.linguistics.ucla.edu/people/hayes/205/readings/ryan_bole_handout.pdf

Selkirk, Elizabeth (1984). On the major class features and syllable theory. In M. Aronoff & R. T.Oehrle (Eds.) *Language sound structure: Studies in phonology presented to Morris Halle by his teacher and students*, (pp. 107-136). Cambridge, MA, London: The MIT Press.

Smolensky, Paul, and Géraldine Legendre (2006). *The harmonic mind: from neural computation to Optimality-theoretic grammar*. Cambridge: MIT Press.

Steriade, Donca (1982). Greek prosodies and the nature of syllabification. PhD dissertation, MIT, Cambridge, Massachusetts.

Wonderly, William L. (1951) Zoque III: morphological classes, affix list, and verbs. *International Journal of American Linguistics* 17: 137-162