

A precedence based lexicon
Eric Raimy
University of Wisconsin, Madison
MOT, March 3, 2007
University of Ottawa

The blind men and the elephant- John Godfrey Saxe (1878)

*It was six men of Indostan,
To learning much inclined,
Who went to see the Elephant
(Though all of them were blind),
That each by observation
Might satisfy his mind.*

1. Design specifications for the lexicon

- Strong Phonology Hypothesis (Frost 1998, 2004, Poeppel, Idsardi and van Wassenhove in press)
“phonological representations are the principle building blocks of the lexicon” (Frost 2004:174)
“the theory of lexical representation that we adopt ... assumes that words are represented in the mental lexicon in terms sequences of discrete segments composed of distinctive features” (Poeppel, Idsardi and van Wassenhove in press :2)
- Storage of idiosyncratic linguistic information
“One approach is to say that the lexicon contains only the information that is completely idiosyncratic” Spencer (1991:47)
-phoneme sequences of morphemes
-lexical strata (Lightner 1972 for Russian, Vance 1987 for Japanese, etc.)
- Economic storage (Yang 2005, Halle 1985, Chomsky 1975)
“...grammar as a data compression device: it strives to minimize the structural descriptions of the relevant linguistic data” Yang (2005:275)
- Frequency effects (Murray and Forster 2004)
“lexical access time changes as a function of the log of frequency, not frequency itself” (Murray and Forster 2004:722)
-age of acquisition and neighborhood effects
- Priming effects (Boudelaa and Marslen Wilson 2000, 2004, Kinoshita and Lupker 2004)
-phonological, semantic, roots, templates
- Uniqueness point effects (Grosjean 1980, Cutler 1995)
-only word that begins with [spɪg] is *spigot* (Cutler 1995:101)
- Growth and acquisition of the lexicon
-Successive Binary Algorithm (Dresher to appear)

*The First approach'd the Elephant,
 And happening to fall
 Against his broad and sturdy side,
 At once began to bawl:
 "God bless me! but the Elephant
 Is very like a wall!"*

2. Building a precedence based lexicon

- What is the most economical way we can store the 7 words in (1)? *recycle*
- Raimy (1999, 2000) proposes that precedence relations are primitives of phonological representations
- McClory and Raimy (2007) proposes that precedence relations contain additional information
 - allows us to reuse phonemes in a very economical way
 - also allows for diacritics for strata (Lightner 1972) and exceptions (Halle and Nevins 2006) to be placed somewhere in the representation

(1) Mini lexicon of NR

	English	Phonology	Semantics	WordCount.com Frequency
a.	train	treXn	locomotive	1276
b.	cap	kAp	hat	4114
c.	blue	bluX	color	973
d.	train	treXn	exercise	1276
e.	truck	tr^k	vehicle	6325
f.	white	wayt	color	363
g.	whitecap	waytkAp	water	not found

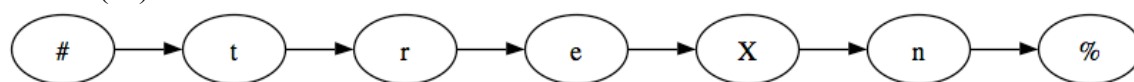
(2) Individual entries

	PBP representations
a.	# → t → r → e → X → n → %
b.	# → k → A → p → %
c.	# → b → l → u → X → %
d.	# → t → r → e → X → n → %
e.	# → t → r → ^ → k → %
f.	# → w → a → y → t → %
g.	# → w → a → y → t → k → A → p → %

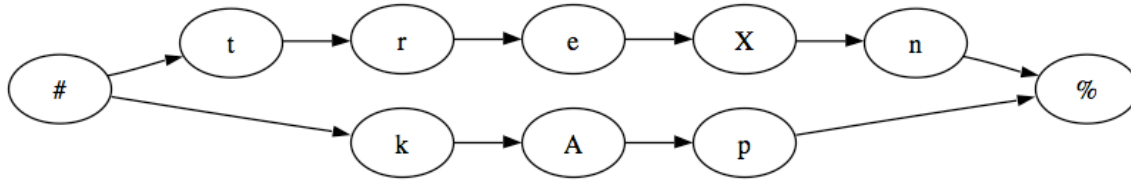
- (2) is not economized at all
- (3) demonstrates how we can reuse phonemes to create a very economic lexicon

(3) An economic lexicon

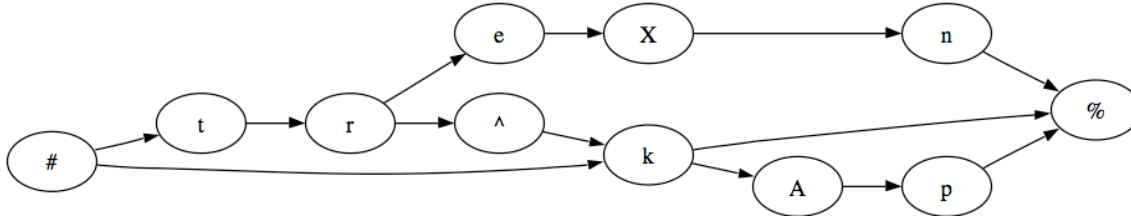
a. train (2a)



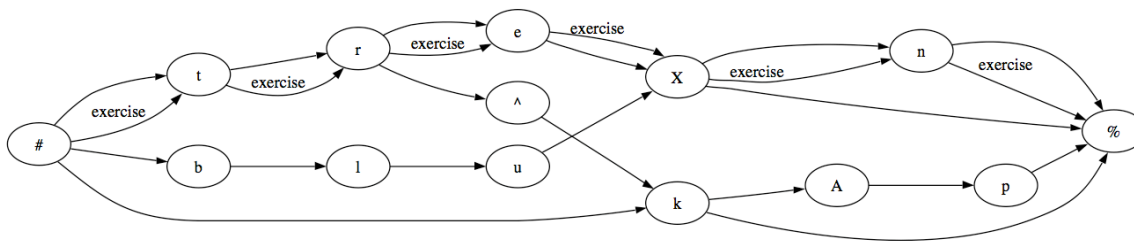
b. train (2a) and cap (2b) *not economic yet*



c. train (2a), cap (2b) and truck (2c) *a bit of economy*

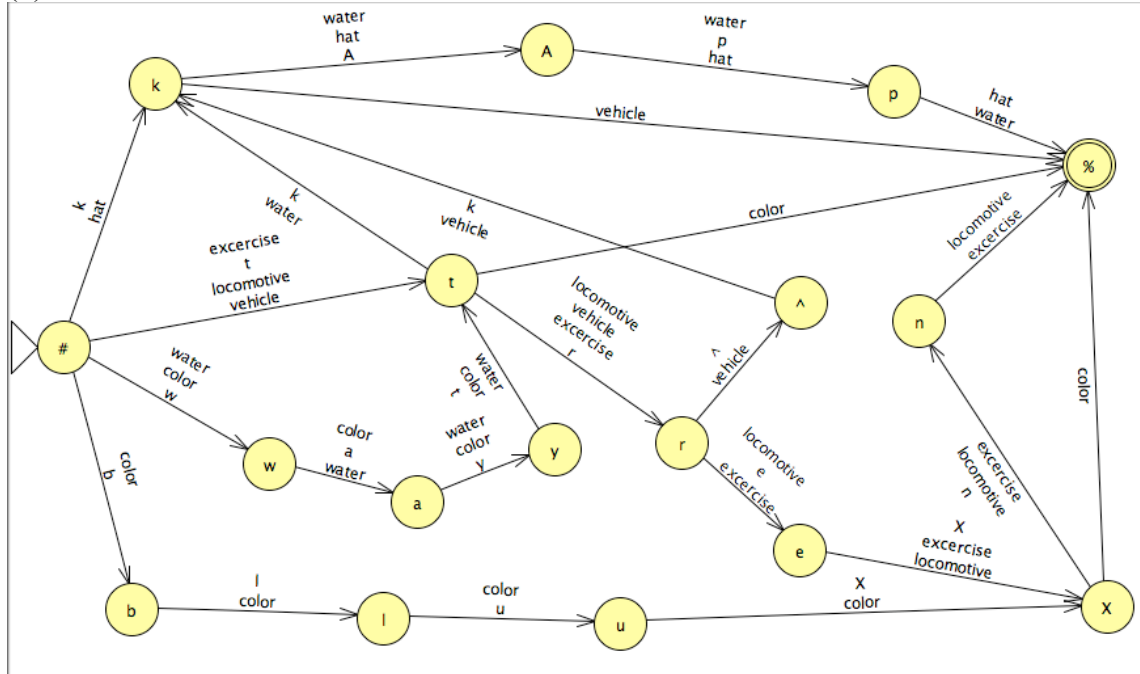


d. train (2a), cap (2b), truck (2c) and train (2d) *good economy here*



- A full mini-lexicon of the forms in (1) is presented in (4) using JFLAP
- JFLAP does ‘maximal compression’ in that individual transitions between phonemes are graphically collapsed
- Compounds follow proposal in Idsardi and Shorey (2007)

(4) JFLAP instantiation of a PBP Lexicon



*The Second, feeling of the tusk,
 Cried, -"Ho! what have we here
 So very round and smooth and sharp?
 To me 'tis mighty clear,
 This wonder of an Elephant
 Is very like a spear!"*

3.0 Useful structural aspects of a PBP lexicon

- The compression in (4) is useful at a certain level because it allows us to see *relatedness between words* that is the basis of our understanding of priming
 - ‘train’ primes ‘train’ whether it is the same semantics or not because of the phonological overlap
 - ‘blue’ primes ‘white’ because of the semantic overlap
- *neighborhood density* can also be seen in the lexicon via number of labels on arcs
 - ‘truck’ in a higher density neighborhood than ‘blue’
- *uniqueness points* can be identified when there is an unambiguous (only a single semantic label) path to %
 - no uniqueness point on ‘train’ because of homophones
 - immediate uniqueness point on ‘b’ because only ‘blue’ in lexicon
 - [tr^] is uniqueness point for ‘truck’

The Third approach'd the animal,
 And happening to take
 The squirming trunk within his hands,
 Thus boldly up and spake:
 "I see," -quoth he- "the Elephant
 Is very like a snake!"

4.0 Statistics on the lexicon

- Can compute statistics over links between phonemes
- Prior information for Bayesian inference (Idsardi 2006) can be extracted from the lexicon
 - what is the distribution of word final segments, i.e. how many paths to ‘%’
- ‘meta-linguistic’ intuitions can be culled from the lexicon
 - what is the ‘functional load’ of a segment? i.e. number of precedence relations to and from a segment
 - /b/ has ‘low functional load’ a single arc in and a single arc out
 - /k/ has ‘high functional load’ three arcs in and two arcs out
- *phonesthemes* are statistical effects in the lexicon based on correlation between paths between segments and semantic marking

*The Fourth reached out an eager hand,
 And felt about the knee:
 "What most this wondrous beast is like
 Is mighty plain," -quoth he,-
 "'Tis clear enough the Elephant
 Is very like a tree!"*

5.0 Another structural view of a PBP lexicon

- Directed graphs are abstract data structures (Aho, Hopcroft and Ullman 1985) in that they can be represented in many different ways
 - adjacency matrix- Heinz (2007)
 - edges and vertices- Idsardi and Shorey (2007)
 - stacks- Raimy (1999)
- Stack based approach to lexicon adds structure to the lexicon in (4) because precedence links must be ordered in the stacks
- Order in stacks can provide rank to produce effects of the *rank hypothesis* of frequency effects Murray and Forster (2004)
- Must search phonological representations (Reiss 2007)

(5) Stack based lexicon

ð-the 1
ə- of 2
æ- and 3
t -to 4
ɪ-in 5
...
t- trade 471	u- 'to' 471	...
...
t-train 1276	r- train 1276	e- train 1276
...
t- tank 2922	æ- tank 2922	...
...	...	e- trail 5638
t-trail 5638	r- trail 5638	...
...
t-tomato 8732	ɛ- tomato 8732	...
...
t- tautology 53276	a- tautology 53276	...
...
#	t	r

*The Fifth, who chanced to touch the ear,
Said- "E'en the blindest man
Can tell what this resembles most;
Deny the fact who can,
This marvel of an Elephant
Is very like a fan!"*

6.0 Multiple levels of representations

- Want to embrace multiple levels of representation in this approach to the lexicon
- Need distinct representations to employ the *entry opening* (Forster, Mohan and Hector 2004) approach to priming
 - project a representation from the lexicon to working memory
 - cull a single representation from stacks of lexicon to edge/vertices representation (Idsardi and Shorey 2007) for phonology

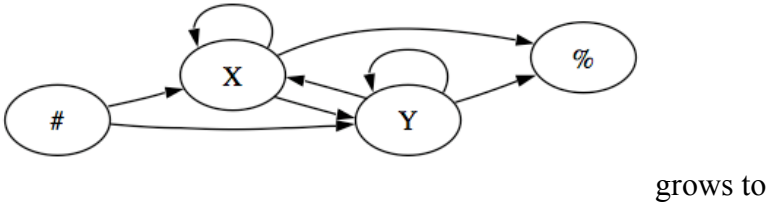
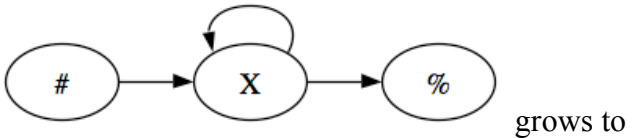
- Converting representations from perception to lexicon also has benefit
 - specific speaker priming (Lively, Pisoni and Goldinger 1994) does not entail more phonetic detail in lexicon
 - priming effect is speed gained in ‘converting to phoneme’

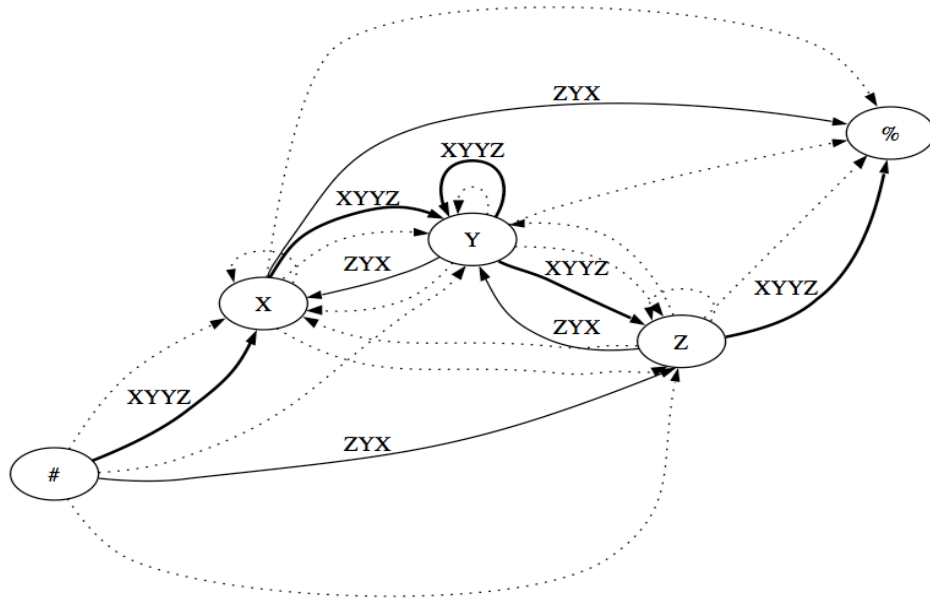
*The Sixth no sooner had begun
 About the beast to grope,
 Then, seizing on the swinging tail
 That fell within his scope,
 "I see," -quoth he,- "the Elephant
 Is very like a rope!"*

7.0 The dynamic lexicon

- Use Drescher (to appear) Successive Binary Division algorithm with a cloning type device to create a truly generative lexicon
- Lexicon starts as a single category and grows in two ways
 - category is split with links cloned to support contrasts- lexicon
 - links for paths through lexicon are cloned and marked with morphemic, semantic, strata, exception marks etc.
- Forces an economized (Idsardi and Raimy to appear, Zuraw 2002) and alphabetic lexicon (see Gagnon 2007 for discussion of economy aspect though)
- Adding a word to the lexicon is also a cloning process
 - clone ‘blank phonological link’ by adding directly above blank (Idsardi and Shorey 2007)
- This approach will produce ‘age of acquisition’ effects because newer words will always be inserted below older words
- Update rank ordering of items based on the distinctive features
 - will produce ‘natural classes’ around lexical neighborhoods giving low frequency words a partial free ride (similar to Yang 2002 approach to irregular verbs in English)
 - explains why low frequency words in dense neighborhoods show faster lexical access than low frequency words in sparse neighborhoods (Andrews 1989, Lively et al 1994)
- Nonce words are processed slowest because the blank phonological links which allow them to be parsed will always be at the ‘bottom of the stack’

(6) A growing lexicon





*And so these men of Indostan
 Disputed loud and long,
 Each in his own opinion
 Exceeding stiff and strong,
 Though each was partly in the right,
 And all were in the wrong!*

MORAL,

*So, oft in theologic wars
 The disputants, I ween,
 Rail on in utter ignorance
 Of what each other mean;
 And prate about an Elephant
 Not one of them has seen!*

8.0 Conclusions

- Precedence based phonology provides a novel way of investigating the structure of the lexicon
- Very close ties between phonology, acquisition and processing in this model
- Assuming parallel searches on lexicon so question of similarity/differences with other lexical access models (see Lively et al 1994 for discussion of following models)
 - logogen models
 - TRACE
 - LAFS
- How is prosody represented in the lexicon?
 - parallel ‘syllable lexicon’?
 - computing prosody over links in lexicon as error correction?
 - Anchor Point calculations on the lexicon? Needed to generalize ‘non-concatenative morphology’ (see Raimy 2007 for discussion)

References

- Aho, Alfred, John E. Hopcroft and Jeffrey D. Ullman (1985) *Data structures and algorithms*. Addison-Wesley Publishing Company.
- Andrews, S. (1989) Frequency and neighborhood effects on lexical access: Activation or search. *Journal of Experimental Psychology: Learning, Memory and Cognition* 15:802-814.
- Boudelaa, Sami and William D. Marslen-Wilson (2004) Abstract morphemes and lexical representation: the CV-skeleton in Arabic. *Cognition* 92:271-303.
- Boudelaa, Sami and William D. Marslen-Wilson (2000) Non-concatenative morphemes in language processing: evidence from Modern Standard Arabic. *Proceedings of the Workshop on Spoken Word Access Processes*. Nijmegen:Max Planck Institute for Psycholinguistics. pp. 23-26.
- Chomsky, Noam (1975) *The Logical Structure of Linguistic Theory*. Plenum Press.
- Cutler, Anne (1995) Spoken word recognition and production. In Joanne L. Miller and Peter D. Eimas *Speech, language and communication, 2nd ed.* Academic Press. pp. 97-137.
- Dresher, E. Elan (to appear) On the Acquisition of phonological contrasts. In *The proceedings of GALA*.
<http://www.chass.utoronto.ca/~dresher/GALA.pdf>
- Forster, Kenneth I. (1976) Accessing the mental lexicon. In R. J. Wales and E. Walker (eds.) *New approaches to language mechanisms*. Amsterdam:North-Holland. pp. 257-287.
- Forster, Kenneth I. (1992) Memory-addressing mechanisms and lexical access. In R. Frost and L. Katz (eds.) *Orthography, phonology, morphology, and meaning*. Amsterdam:North-Holland. pp. 413-434.
- Forster, Kenneth I., Kathleen Mohan and Jo Hector (2004) The mechanisms of masked priming. In Sachiko Kinoshita and Stephen J. Lupker (eds.) *Masaked Priming: the state of the art*. Psychol. Press. pp. 3-38.
- Frost, Ram (1998) Towards a strong phonological theory of visual word recognition: True issues and false trails. *Psychological Bulletin* 123:71-99.
- Frost, Ram (2004) The robustness of phonological effects in fast priming. In Sachiko Kinoshita and Stephen J. Lupker (eds.) *Masaked Priming: the state of the art*. Psychol. Press. pp. 173-192.
- Gagnon, Michaël (2007) Token vs. Type identity. Talk presented at the CUNY Conference on Precedence Relations.
<http://www.cunyphonologyforum.net/forum.php>
- Grosjean, F. (1980) Spoken word recognition processes and the gating paradigm. *Perception & Psychophysics* 28:299-310.
- Lightner, Theodor M. (1972) *Problems in the theory of phonology, vol. 1, Russian phonology and Turkish phonology*. Linguistic Research, Inc.
- Lively, Scott E., David B. Pisoni and Sephen D. Goldinger (1994) Spoken word recognition: research and theory. In Morton Ann Gernsbacher (ed) *Handbook of psycholinguistics* Academic Press. pp.265-302.
- Halle, Morris (1985) Speculations about the representations of words in memory. In Morris Halle (2002) *From memory to speech and back*. Mouton de Gruyter pp.122-136.
- Halle, Morris and Andrew Nevins (2006) Rule application in phonological theory. To appear in Eric Raimy and Charles Cairns (eds.) *Contemporary Views on Architecture and Representations in Phonological Theory*. MIT Press.
- Heinz, Jeffrey (2007) Learning long-distance agreement phonotactics. Handout from Talk given at the 81st Annual LSA Meeting, Anaheim, CA.
- Idsardi, William J. (2006) A Bayesian approach to loanword adaptation. Poster presented at the 80th Annual LSA Meeting, Albuquerque, NM.
- Idsardi, William and Eric Raimy (to appear) Reduplicative Economy. In Andrew Nevins and Bert Vaux. *Rules, constraints and phonological phenomena*. Oxford Univeristy Press. <http://ling.umd.edu/~idsardi/work/2002redecon.pdf>
- Idsardi, William J. and Rachel Shorey (2007) Unwinding morphology. Talk presented at the CUNY Conference on Precedence Relations.
<http://www.cunyphonologyforum.net/forum.php>
- JFLAP 6.0 <http://www.jflap.org/>
- Kinoshita, Sachiko and Stephen J. Lupker (2004) *Masked priming: The state of the art* Psychology Press.
- McClory, Daniel and Eric Raimy (2007) Enhanced Edges: morphological influence on linearization. Poster presented at The 81st Annual Meeting of the Linguistics Society of America. Anaheim, CA.
- Murray, W. S. and Kenneth I. Forster (2004) Serial mechanisms in lexical access: the Rank Hypothesis. *Psychological Review* 111(3):721-756.
- Poeppl, David, William J. Idsardi and Virginie van Wassenhove (in press) Speech perception at the interface between neurobiology and linguistics. *Proc. Royal Soc. London*
- Raimy, Eric (2007) Precedence theory, root and template morphology, priming effects and the structure of the lexicon. Talk presented at the CUNY Conference on Precedence Relations. <http://www.cunyphonologyforum.net/forum.php>
- Raimy, Eric (2000) *The phonology and morphology of reduplication*. Mouton de Gruyter.
- Raimy, Eric (1999) Representing reduplication. Ph.D. dissertation, University of Delaware.
- Reiss, Charles (2007) Adjacency as a long distance relation. Talk presented at the CUNY Conference on Precedence Relations.
<http://www.cunyphonologyforum.net/forum.php>
- Saxe, John Godfrey (1878) The blind men and the elephant. In Williams James Spencer (ed) *Poetry of America* Convent Garden. pp. 150-152. Available from books.google.com
- Spencer, Andrew (1991) *Morphological Theory*. Blackwell Publishers.
- Vance, Timothy (1987) *An introduction to Japanese phonology*. State University of New York Press.
- Yang, Charles (2002) *Knowledge and learning in natural language*. Oxford University Press.
- Yang, Charles (2005) On productivity. In *Linguistic Variation Yearbook* 5. John Benjamins Publishing Co. pp. 265-302.
- Zuraw, Kie (2002) Aggressive reduplication. *Phonology* 19:395-439.