Blake Allen, b.allen@alumni.ubc.ca
Michael Becker, michael.becker@stonybrook.edu

Phonology 2014 @ MIT
September 19–21, 2014

Learning alternations from surface forms with sublexical phonology*

Highlights:

- The generative approach to learning alternations relies on finding URs for morphemes.

  Algorithmic approaches to finding URs haven't been able to handle data that is realistic in terms of size and variation/noise.

- The sublexical approach (Becker & Gouskova 2013): Don't search for URs.

  Partition the lexicon by morphological operation, learn a MaxEnt grammar for each sublexicon.

- We present a learner that partitions the lexicon in the desired way by creating a flurry of hypotheses and subsequent reduction into sublexicons.

- Nonce words are derived via sublexicon application, creating a set of derivatives with a probability distribution over them.

1 UR for the English plural suffix

(1) General agreement about operations and their distribution:

- "add [-ɪz]" after stridents
- "add [-s]" after voiceless sounds (except stridents)
- "add [-z]" after voiced sounds (except stridents)

* For their thoughtful comments and suggestions, we thank Adam Albright, Bruce Hayes, Colin Wilson, and Kie Zuraw. Special thanks to Maria Gouskova, whose thinking permeates much of this work. We also thank Peter Kleiweg for sharing the code for the minimal edit distance algorithm, and Ilia Kurenkov and Omer Preminger for help with the Javascript implementation. Remaining errors are the other author's.

(2) Traditional UR-based analysis:

- Insightfully and elegantly choose [z] as the UR /z/
- Grammar:

/tɹɪk + z/	OCP (strident)	Dep	Agree (voice)	Ident (voice)
a. tɹɪkz			*!	
b. ☞ tɹɪks				*
c. tɹɪkɪz		*!		

/kiʃ + z/	OCP (strident)	Dep	Agree (voice)	Ident (voice)
a. kiʃz	*!	*		
b. kiʃs	*!			*
c. ☞ kiʃɪz		*		

(3) Choosing among the surface forms works here; not a general solution.

(4) Algorithmic approaches to finding URs:

- Tesar et al. (2003) et seq.: "surgery" combines information from the surface forms to make a composite UR.
- Jarosz (2006): generate URs from the rich base to maximize the likelihood of the lexicon.

Not general solutions (e.g., floating elements, subcategorization).
Unknown whether these approaches generalize to realistic data.

(5) Skepticism about the search for URs:
Working "inside-out" = the base is the root's UR (Hayes 1995, 1999b; Becker 2009), limits the search for URs to affixes.

(6) No URs at all:
Minimal Generalization Learner (MGL, Albright & Hayes 2002, 2003b, 2006).
On the joys of not having URs: Albright (2002a,b, 2006, 2008a,b).

(7) No *known* way to find URs algorithmically for realistic data.
UR discovery is an unsolved problem, perhaps unsolvable.

The sublexical analysis (Becker & Gouskova 2013):

(8) Partition the lexicon by morphological operation, make a gatekeeper (phonotactic MaxEnt grammar) for each sublexicon:

- The "add [ɪz]" words + heavily weighted *[−strident]#
- The "add [s]" words + heavily weighted *[+voice]#
- The "add [z]" words + heavily weighted *[−voice]#

(9) Sublexicon = list of paradigms + uniform morphological operation + MaxEnt grammar (or pair of gatekeeper grammar and grammar proper).

(10) Generalization to nonce words via application of each sublexicon.
Three sublexicons → three derivatives:

- [wʌgɪz] violates *[−strident]#
- [wʌgs] violates *[+voice]#
- [wʌgz] is well-formed, gets most of the probability.

(11) Gatekeeper grammars assess bases only; no role for OCP(strident), Ident(voice).

2 Finding the English plural affix

(12) Minimum edit distance alignment (Levenshtein algorithm, Needleman–Wunsch algorithm):

b	ɹ	ʌ	ʃ	∅	∅
b	ɹ	ʌ	ʃ	ɪ	z

Code based on Peter Kleiweg's `http://www.let.rug.nl/~kleiweg/lev/`

(13) Augmented with featural similarity (cf. Nerbonne & Heeringa 2010; Beijering et al. 2008; Spruit et al. 2007; Heeringa 2004; Nerbonne et al. 1996)

Challenge I: finding the position of affixes is surprisingly hard

(14) Ulwa (Hale & Blanco 1989; Bromberger & Halle 1988; McCarthy & Prince 1993b,a): left-oriented *ka* sometimes appears at the right edge

base	possessed	
bas	bas.*ka*	'hair'
sa.paa	sa.paa.*ka*	'forehead'
suu.lu	suu.*ka*.lu	'dog'
kuh.bil	kuh.*ka*.bil	'knife'

How do you know that the [ɪz] of [bɹʌʃɪz] is a right-oriented affix?

Challenge II: affixes hard to identify when they share segments with the base

(15) English: what is the analysis of [ɹoʊz ∼ ɹoʊzɪz] 'rose'?

ɹ	oʊ	z	∅	∅
ɹ	oʊ	z	ɪ	z

vs.

ɹ	oʊ	∅	∅	z
ɹ	oʊ	z	ɪ	z

The alignment algorithm greedily matches edge segments.
Even worse in Russian [l^{j}ef ∼ l^{j}vof] 'lion':

l^j	e	f	∅	∅
l^j	∅	v	o	f

vs.

l^j	∅	e	f
l^j	v	o	f

Solution: generate many hypotheses about the affix, consume the smaller ones.

(16) Hypotheses generated from 'kiss', 'brush', 'rose':

[kɪs ∼ kɪsɪz] →

k	ɪ	s	∅	∅
k	ɪ	s	ɪ	z

→ 'add [ɪz] at right edge'
→ 'add [ɪz] after 3^{rd} seg'

[bɹʌʃ ∼ bɹʌʃɪz] →

b	ɹ	ʌ	ʃ	∅	∅
b	ɹ	ʌ	ʃ	ɪ	z

→ 'add [ɪz] at right edge'
→ 'add [ɪz] after 4^{th} seg'

[ɹoʊz ∼ ɹoʊzɪz] →

ɹ	oʊ	∅	∅	z
ɹ	oʊ	z	ɪ	z

→ 'add [zɪ] before last seg'
→ 'add [zɪ] after 2^{nd} seg'

(17) Hypothesis reduction (simplified, see complete version in 47):

Hypothesis H_L is consumed by hypothesis H_W iff
H_W and H_L are *idempotent* on H_L's bases
and H_W covers *more lexical items* than H_L.

(18) Hypothesis reduction eliminates all but "add [ɪz] at right edge":
"add [ɪz] at right edge" consumes "add [ɪz] after third segment"
"add [ɪz] at right edge" consumes "add [zɪ] before final segment"

(19) Together with the non-strident-final words, the learner finds the three desired sublexicons:

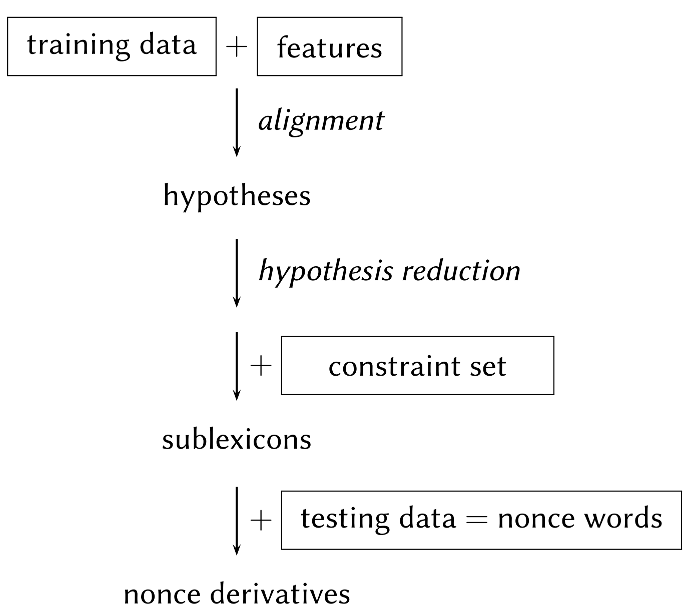
- "add [-ɪz] at right edge"
- "add [-s] at right edge"
- "add [-z] at right edge"

(20) Once hypothesis reduction is complete, MaxEnt grammars are fitted for each sublexicon.

(21) We use our own Javascript MaxEnt module with support for priors etc.

(22) Application to nonce words: [wʌg] → [wʌgɪz], [wʌgs], [wʌgz]

(23) Summary:



3 Realistic English plurals

Becker, Nevins & Levine (2012): data for real and nonce plurals

(24) Irregular voicing alternations with noun-final [f/θ]

	f			θ		
stressed, heavy	loʊf	loʊvz	‘loaf’	maʊθ	maʊðz	‘mouth’
	bɹif	bɹifs	‘brief’	feɪθ	feɪθs	‘faith’
stressed, light	hʊf	hʊvz	‘hoof’	bɚθ	bɚðz	‘berth’
	blʌf	blʌfs	‘bluff’	bɹɛθ	bɹɛθs	‘breath’
unstressed	ʃɛɹɪf	ʃɛɹɪfs	‘sheriff’	mæmɪθ	mæmɪθs	‘mammoth’

Tons of variation (giraffe, chief, paragraph, roof, truth, path, youth, etc.)

(25) Predictors (in real words and in the treatment of nonce words):

- Place: voicing more likely with [f] than with [θ]
- Weight: voicing more likely following long vowel/diphthong/coda C
- Stress: voicing more likely following a stressed vowel

(26) Our goal: train the model on real words, match participants’ judgments of nonce words.

Finding the best sublexicons:

(27) Faithful plurals, e.g. [bɹɛθ ∼ bɹɛθs], go in the “add [s]” sublexicon.

(28) **Segmental** generalization for alternators:

- “add [z] after final segment, change final segment from [f] to [v]”
- “add [z] after final segment, change final segment from [θ] to [ð]”

(29) **Featural** generalization for alternators (product-oriented):

- “add [z] after final segment, make final segment [+voice]”

(30) Fewer sublexicons = bigger sublexicons = potential for broader generalizations.

Running the learner:

(31) Training data:
10,000 most frequent nouns from CELEX + their faithful/regular plurals.
126 f/θ-final existing words from Becker et al. (2012), each with faithful plural + probability and with voiced plural + probability.

(32) We set a productivity threshold at 10. (to do: verify empirically)

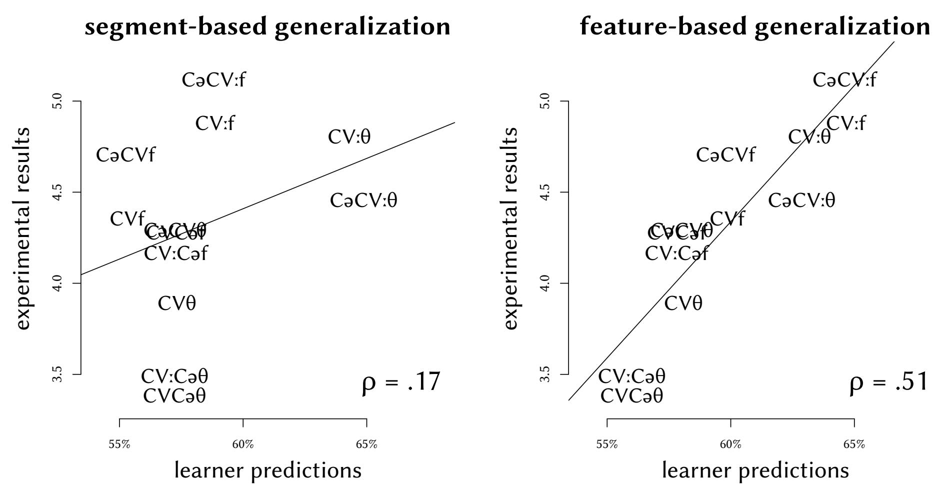
(33) Constraints (excluding zero-weighted):

*[+strident]#	*(f|θ)#
*[0strident]#	*θ#
*[−strident]#	*(¬f|θ)#
*[+voice]#	*[−syll, +son](f|θ)#
*[−voice]#	*[+syll, +stress](f|θ)#
*[+syllabic]#	*[+syll, +long](f|θ)#
*[−syllabic]#	

Supplied by the analyst (cf. induction in Moore-Cantwell & Staubs 2014)
Same constraints (+priors) used in all sublexicons.

(34) Testing: 132 singular nonce words from Becker et al. (2012).

(35) Our predictions vs. the experimental results:



(36) Improved prediction with feature-based generalizations, which put all the alternators in the same sublexicon.

Is there a UR-based analysis of English that includes the alternating plurals?

(37) Perhaps some kind of underspecification?

	UR	singular	plural	
voiced	/lʌv/	lʌv	lʌvz	'love'
alternating	/ʃɛlF/	ʃɛlf	ʃɛlvz	'shelf'
voiceless	/læf/	læf	læfs	'laugh'

Getting the grammatical and ungrammatical patterns is not enough.

The analysis must also generalize the place, weight, and stress factors from the lexicon to nonce words.

The analysis needs to be learnable.

(38) Traditionally, UR-based analyses try to cover multiple affixes.
These analyses will have too many types for underspecification.

	UR?	singular	plural	verb	genitive	
voiced	/lʌv/	lʌv	lʌvz	lʌv	lʌvz	'love'
alt. pl+V	/ʃɛlF/	ʃɛlf	ʃɛlvz	ʃɛlv	ʃɛlfs	'shelf'
alt. pl	/naɪF/	naɪf	naɪvz	naɪf	naɪfs	'knife'
alt. V	/bəliF/	bəlif	bəlifs	bəliv	bəlifs	'belief'
voiceless	/læf/	læf	læfs	læf	læfs	'laugh'

Not clear what the UR-based analysis/single grammar analysis would be.

(39) Finding the analysis using insight $\neq$ there is a learner than can find it.

(40) The sublexical analysis does not attempt to learn more than one morphological category at a time.

(41) Similarities among affixes (e.g. lexical conservatism, Steriade 1999) = additional information that needs to be learned (see, e.g., Albright 2010).

4 More on finding sublexicons

4.1 Infixation: multiple reduction

No single hypothesis available to subsume undesired hypotheses:

(42) Tagalog [in]-infixation (Zuraw 2007)

	base	derivative	
V-initial	abot	in-abot	‘attain’
CV-initial	bago	b-in-ago	‘big’
CCV-initial	pɾoblema	p-in-ɾoblema	‘problem’
		pɾ-in-oblema	

(Ignore V-initial words if you believe they are ʔ-initial, as in Halle 2001)

(43) Our learner needs to discover that the affix is left-oriented.

(44) Desirable hypotheses are not supersets of undesirable ones

left-oriented		right-oriented
before first segment	in x x x	**before antepenult**
before first segment	in x x x x	before preantepenult
before first segment	in x x x x x	before prepreantepenult
after first segment	x in x x	before penult
after first segment	x in x x x	**before antepenult**
after first segment	x in x x x x	before preantepenult
after second segment	x x in x	before last
after second segment	x x in x x	before penult
after second segment	x x in x x x	**before antepenult**

e.g. the undesirable **“before antepenultimate segment”** is split between three desirable hypotheses.

(45) Solution: allow multiple reduction; a hypothesis may be consumed by arbitrary combinations of hypotheses.

(46) Increases the search space, but it’s still surprisingly manageable.
Not much of a bottleneck in our implementation relative to single reduction.

We start with reducing the smallest hypotheses, which usually allows us to only explore a small fraction of the worst case scenario.

Hypothesis reduction (complete):

(47) Hypothesis H_L is consumed by set of hypotheses $\{H_{W1}, H_{W2}, ...\}$ iff $\{H_{W1}, H_{W2}, ...\}$ and H_L are idempotent on H_L's bases, and $\{H_{W1}, H_{W2}, ...\}$ cover more lexical items than H_L.

(48) The algorithm we implemented:
For each base in each hypothesis H_L, generate derivatives by applying the operations from all other hypotheses.
If the resulting large set of predicted derivatives is a superset of H_L's observed derivatives, consume H_L, and assign each base in H_L to the biggest hypothesis that can generate its derivative.

(49) Find a link to the source code (on Github) from the learner's website
`http://sublexical.phonologist.org/`

Final analysis of Tagalog:

(50) Three remaining sublexicons after hypothesis reduction:

- "add [in] before first segment"
- "add [in] after first segment"
- "add [in] after second segment"

Constraints on sonority sequences ensure the intended distribution.

(51) No formal expression for the unity of [in] in the three sublexicon.
Is there evidence that this is a problem?

(52) Separate sublexicons are need when an affix has multiple shapes and/or multiple positions.
In Tagalog, the shape is constant, but the position is not.

Moore-Cantwell & Staubs (2014):

(53) Inspired by Becker & Gouskova (2013); Gouskova & Newlin-Łukowicz (2013)

- Hypothesis reduction ("bundle merger") is random; single reduction only.
- Hypothesis reduction interleaved with markedness constraint induction (cf. Hayes & Wilson 2008).
- Operational constraints (cf. anti-correspondence, Hayes 1999a,b; Albright & Hayes 2006).
- Single grammar for the language with indexed constraints — almost certainly a notational variant of our one grammar per sublexicon.

4.2 Non-concatenative morphology: metathesis

Rotuman (McCarthy 2000, based on Churchward 1940)

(54) General verb-final CV → VC metathesis.
Fusion if final V is front and not lower than penult V.

a.	hosa	hoas	'flower'
	pure	puer	'to rule'
b.	hoti	høt	'to embark'
	futi	fyt	'to pull'

(55) Sublexical analysis:

- "metathesize the last two segments"
- "front the antepenultimate segment and delete the final segment"

Constraints on vowel quality regulate the choice of sublexicon.

(56) Our learner encodes a genuine metathesis operation; expected to apply beyond the training space (cf. Berent, Wilson, Marcus & Bemis 2012)

(57) Fronting + deletion in [futi ∼ fyt], not metathesis.
Any reason to think that this is a problem?

4.3 Supra-segmental loci

Russian vowel deletion (Gouskova & Becker 2013; Becker & Gouskova 2013):

(58) Mid vowel deletion possible with simple coda; complex coda blocks

	nominative	genitive	
simple coda	p^jos	ps-a	'dog'
	kast^{j}or	kastr-a	'fire'
complex coda	as^{j}otr	*astr-a	'sturgeon'

(59) Possible generalizations:

- Delete final vowel, add [a] at right edge
- Delete penultimate segment, add [a] at right edge

Equally good coverage of the data.

(60) Differential application to final complex codas:

- as^{j}otr ∼ *astr-a (delete final vowel)
- as^{j}otr ∼ ***as^{j}or-a (delete penult segment)

Both bad, but [as^{j}otr ∼ as^{j}or-a] is much worse.

(61) Our solution:

- Designate certain segments as privileged positions; currently defined as [+syllabic].
- Generate hypotheses that refer to these positions, e.g. "delete final nucleus".
- Supra-segmental hypotheses are preferred, ceteris paribus.

  "Delete final nucleus" and "delete penultimate segment" are equally supported by the Russian data; an inherent bias is needed.

Future work: extend the range of supra-segmental positions to allow fuller coverage of infixation, stress shifts, templatic morphology(!), etc.

5 Comparison with the MGL

Minimal Generalization Learner (Albright & Hayes 2002, 2003b, 2006)

(62) Build a rule from each paradigm, then generalize:

kiʃ ∼ kiʃɪz	∅ → [ɪz] / k i ʃ __
ɹoʊz ∼ ɹoʊzɪz	∅ → [ɪz] / ɹ oʊ z __
Generalization:	∅ → [ɪz] / X [+strident +cont] __

(63) Similarities: Both learners—

- Learn from realistic data, in terms of quantity and irregularity.
- Learn many generalizations for a given pair of morphological categories.
- Generate derivatives for nonce words + likelihood.

(64) MGL is rule-based
Rules combine operation + environment (necessarily local, though see Albright & Hayes 2003a) + reliability/confidence.

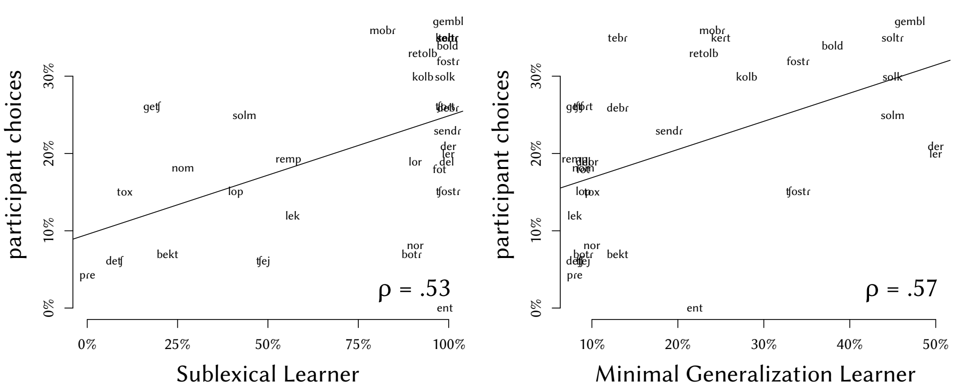
(65) The Sublexical Learner is constraint-based
The grammar controls the distribution of operations and the probability of derivatives.
Allows the exploration of non-local and supra-segmental environments.

Testing on Spanish (lexical data and wugs from Albright et al. 2001):

(66) Diphthongization in Spanish verbs, e.g., [kont- ∼ kwent-] ‘count’.

(67) Our constraint set based on the environments found by the MGL.

(68) The two learners generate similar results in the wug test.



6 Future direction: crossed irregularity

Turkish possessive (Becker, Ketrez & Nevins 2011):

(69) Irregular voicing alternations and deletion

	stem	possessive	
faithful	halk	halk-V	‘people’
voicing	renk	reng-V	‘color’
deletion	gøk	gø-V	‘sky’

(70) Irregular choice of the possessive vowel (Clements & Sezer 1982)

	stem	possessive	
harmonic	ʤennet	ʤennet-i	‘paradise’
	bojut	bojut-u	‘dimension’
	kykyrt	kykyrt-y	‘sulphur’
	teferruat	teferruat-ɨ	‘detail’
disharmonic	sejahat	sejahat-i	‘trip’

Voicing alternations and vowel (dis-)harmony occur in the same words; can they be learned separately?

(71) When given both patterns, our learner makes too many sublexicons:

- “add [i]”
- “voice final segment, add [i]”
- “deleted final segment, add [i]”
- “add [u]”
- “voice final segment, add [u]”
- “delete final segment, add [u]”
- etc. for each combination

Possible solution: learn each operation separately, but also learn to what extent operations coöcurr.

7 Conclusions

One grammar or more?

(72) Learning morphophonology with one grammar:

- Tesar & Smolensky (1998, 2000); Tesar et al. (2003)
- Jarosz (2006)
- Riggle (2006)
- Rasin & Katzir (2013)

(73) But languages are not uniform, even when only considering regular morphophonology:

- Bases are phonologically different from derivatives, e.g. English [fs]#
- Affixes are phonologically different from each other, e.g. English [ɚ]

Cophonologies (Inkelas et al. 1996; Inkelas & Zoll 2007; Anttila 2002, a.o)

Indexation (Pater 2000, 2006, 2008; Fukazawa 1999; Itô & Mester 1999; Kawahara et al. 2002; Flack 2007; Gouskova 2007; Becker 2009; Becker et al. 2011, a.o.)

Learning grammars instead of representations:

(74) English plurals:

- We showed an analysis that learns faithful plurals and voicing alternations from a realistic lexicon (large and exceptionful).
- The analysis learns predictors of (non-)alternation: place, syllable weight, and stress.

  Constraints are supplied by the analyst (for now).
- The model passes a wug test, creating plurals with a probability distribution over then.

  Be a smart consumer! Don't pay for an analysis before you see it generate nonce derivatives.

(75) If there is a single-grammar analysis (UR-based?) of the English plural, nobody has shown how to learn it.

(76) See several other datasets and their analyses at
`http://sublexical.phonologist.org/`
Try it with your data!

References

Albright, Adam (2002a). *The Identification of Bases in Morphological Paradigms*. Ph.D. dissertation, UCLA.

Albright, Adam (2002b). The lexical bases of morphological well-formedness. In sabrina bendjaballah, w. u. dressler, o. e. pfeiffer & m. voeikova (eds.) *Morphology 2000: Selected papers from the 9th Morphology Meeting, Vienna, 24-28 February 2000*, Benjamins. 1–8.

Albright, Adam (2006). Inflectional paradigms have bases too: Arguments from Yiddish. Ms. MIT.

Albright, Adam (2008a). A Restricted Model of UR Discovery: Evidence from Lakhota. Ms. MIT.

Albright, Adam (2008b). Explaining universal tendencies and language particulars in analogical change. In Jeff Good (ed.) *Language Universals and Language Change*, Oxford University Press. 36 pp.

Albright, Adam (2010). Lexical and morphological conditioning of paradigm gaps. In *Modeling Ungrammaticality in Optimality Theory*, London: Equinox.

Albright, Adam, Argelia Andrade & Bruce Hayes (2001). Segmental environments of Spanish diphthongization. In Adam Albright & Taehong Cho (eds.) *UCLA Working Papers in Linguistics 7 (Papers in Phonology 5)*, UCLA. 117–151.

Albright, Adam & Bruce Hayes (2002). Modeling English past tense intuitions with minimal generalization. In Michael Maxwell (ed.) *Proceedings of the sixth meeting of the ACL special interest group in computational phonology*. Philadelphia: ACL, 58–69.

Albright, Adam & Bruce Hayes (2003a). Learning nonlocal environments. Handout from 2003 LSA talk.

Albright, Adam & Bruce Hayes (2003b). Rules vs. Analogy in English past tenses: a computational/experimental study. *Cognition* **90**. 119–161.

Albright, Adam & Bruce Hayes (2006). Modeling productivity with the gradual learning algorithm: The problem of accidentally exceptionless generalizations.

In Gisbert Fanselow, Caroline Féry, Matthias Schlesewsky & Ralf Vogel (eds.) *Gradience in Grammar*, Oxford University Press. 185–204.

Anttila, Arto (2002). Morphologically conditioned phonological alternations. *Natural Language and Linguistic Theory* **20**. 1–42.

Becker, Michael (2009). *Phonological Trends in the Lexicon: The Role of Constraints*. Ph.D. dissertation, University of Massachusetts Amherst.

Becker, Michael & Maria Gouskova (2013). Source-oriented generalizations as grammar inference in Russian vowel deletion. Ms. lingbuzz/001622.

Becker, Michael, Nihan Ketrez & Andrew Nevins (2011). The surfeit of the stimulus: Analytic biases filter lexical statistics in Turkish laryngeal alternations. *Language* **87**. 84–125.

Becker, Michael, Andrew Nevins & Jonathan Levine (2012). Asymmetries in generalizing alternations to and from initial syllables. *Language* **88**. 231–268.

Beijering, Karin, Charlotte Gooskens & Wilbert Heeringa (2008). Predicting intelligibility and perceived linguistic distance by means of the levenshtein algorithm. In *Linguistics in the Netherlands 2008*, John Benjamins.

Berent, Iris, Colin Wilson, Gary Marcus & Doug Bemis (2012). On the role of variables in phonology: Remarks on Hayes and Wilson 2008. *Linguistic Inquiry* **43**. 97–119.

Bromberger, Sylvain & Morris Halle (1988). Conceptual issues in morphology.

Churchward, C. Maxwell (1940). *Rotuman grammar and dictionary*. Sydney [Reprinted by AMS Press, 1978]: Australasia Medical Publishing Co.

Clements, G. N. & Engin Sezer (1982). Vowel and consonant disharmony in Turkish. In *The Structure of Phonological Representations, Part II*, Dordrecht: Foris. 213–255.

Flack, Kathryn (2007). Templatic morphology and indexed markedness constraints. *Linguistic Inquiry* **38**.

Fukazawa, Haruka (1999). *Theoretical implications of OCP effects on features in Optimality Theory*. Ph.D. dissertation, University of Maryland, College Park, MD.

Gouskova, Maria (2007). The reduplicative template in tonkawa. *Phonology* **24**. 367–396.

Gouskova, Maria & Michael Becker (2013). Russian yer alternations are governed by the grammar. *Natural Language and Linguistic Theory* **31**. 735–765.

Gouskova, Maria & Sofya Kasyanenko (2014). Generalization as sublexical phonotactics: A nonce study of Russian allomorphy. Talk given at Phonlogy 2014, MIT.

Gouskova, Maria & Luiza Newlin-Łukowicz (2013). Phonological selectional restrictions as sublexical phonotactics. Ms. NYU.

Hale, Kenneth & Abanel Lacayo Blanco (1989). *Diccionario elemental del Ulwa (Sumu meridional)*. Cambridge, MA: Center for Cognitive Science, MIT.

Halle, Morris (2001). Infixation versus onset metathesis in Tagalog, Chamorro, and Toba Batak. In Kenneth Locke Hale & Michael Kenstowicz (eds.) *Ken Hale: A Life in Language*, MIT Press. 153–168.

Hayes, Bruce (1995). On what to teach the undergraduates: Some changing orthodoxies in phonological theory. In Ik-Hwan Lee (ed.) *Linguistics in the Morning Calm 3*, Seoul: Hanshin. 59–77.

Hayes, Bruce (1999a). Anticorrespondence in Yidiɲ. Ms. UCLA.

Hayes, Bruce (1999b). Phonological restructuring in Yidiɲ and its theoretical consequences. In Ben Hermans & Marc van Oostendorp (eds.) *The derivational residue in phonology*, Amsterdam: Benjamins. 175–205.

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

Heeringa, Wilbert (2004). *Measuring Dialect Pronunciation Differences using Levenshtein Distance*. Ph.D. dissertation, University of Groningen.

Inkelas, Sharon, C. Orhan Orgun & Cheryl Zoll (1996). Exceptions and static phonological patterns: cophonologies vs. prespecification.

Inkelas, Sharon & Cheryl Zoll (2007). Is grammar dependence real? A comparison between cophonological and indexed constraint approaches to morphologically conditioned phonology. *Linguistics* **45**. 133–171.

Itô, Junko & Armin Mester (1999). The phonological lexicon. In Natsuko Tsujimura (ed.) *The Handbook of Japanese Linguistics*, Blackwell. 62–100.

Jarosz, Gaja (2006). *Rich Lexicons and Restrictive Grammars - Maximum Likelihood Learning in Optimality Theory*. Ph.D. dissertation, Johns Hopkins University.

Kawahara, Shigeto, Kohei Nishimura & Hajime Ono (2002). Unveiling the unmarkedness of Sino-Japanese. *Japanese/Korean Linguistics* **12**. 140–151.

McCarthy, John J. (2000). The prosody of phrase in rotuman. *Natural Language and Linguistic Theory* **18**. 147–197.

McCarthy, John J. & Alan Prince (1993a). Generalized alignment. In Geert Booij & Jaap van Marle (eds.) *Yearbook of Morphology*, Dordrecht: Kluwer. 79–153.

McCarthy, John J. & Alan Prince (1993b). *Prosodic morphology I: Constraint interaction and satisfaction.* New Brunswick: Rutgers University Center for Cognitive Science. Available as ROA-482 on the Rutgers Optimality Archive, http://roa.rutgers.edu.

Moore-Cantwell, Claire & Robert Staubs (2014). Modeling morphological subgeneralizations. In John Kingston, Claire Moore-Cantwell, Joe Pater & Robert Staubs (eds.) *Proceedings of Phonology 2013*, Linguistic Society of America.

Nerbonne, John & Wilbert Heeringa (2010). Measuring dialect differences. In Peter Auer & Jürgen Erich Schmidt (eds.) *Language and Space. An international Handbook of Linguistic Variation*, Mouton de Gruyter, vol. 1. 550–567.

Nerbonne, John, Wilbert Heeringa, Eric van den Hout, Peter van de Kooi, Simone Otten & Willem van de Vis (1996). Phonetic distance between dutch dialects. In G. Durieux, W. Daelemans & S. Gillis (eds.) *CLIN VI: Proceedings of the Sixth CLIN Meeting*, Centre for Dutch Language and Speech (UIA). 185–202.

Pater, Joe (2000). Nonuniformity in English secondary stress: The role of ranked and lexically specific constraints. *Phonology* **17**. 237–274.

Pater, Joe (2006). The locus of exceptionality: Morpheme-specific phonology as constraint indexation. In Leah Bateman, Michael O'Keefe, Ehren Reilly & Adam Werle (eds.) *Papers in Optimality Theory III*, Amherst, MA: GLSA. 259–296.

Pater, Joe (2008). Morpheme-specific phonology: Constraint indexation and inconsistency resolution. In Steve Parker (ed.) *Phonological Argumentation: Essays on Evidence and Motivation*, London: Equinox. 123–154.

Rasin, Ezer & Roni Katzir (2013). On evaluation metrics in Optimality Theory. Ms. MIT and TAU.

Riggle, Jason (2006). Using entropy to learn OT grammars from surface forms alone. In Donald Baumer, David Montero & Michael Scanlon (eds.) *Proceedings of WCCFL 25*, Cascadilla Proceedings Project. 346–353.

Spruit, Marco René, Wilbert Heeringa & John Nerbonne (2007). Associations among linguistic levels. *Lingua* **119**. 1624–1642.

Steriade, Donca (1999). Lexical conservatism in French adjectival liaison. In Jean-Marc Authier, Barbara E. Bullock & Lisa A. Reed (eds.) *Formal Perspectives in Romance Linguistics*, John Benjamins. 243–270.

Tesar, Bruce, John Alderete, Graham Horwood, Nazarré Merchant, Koichi Nishitani & Alan Prince (2003). Surgery in language learning. In G. Garding & M. Tsujimura (eds.) *Proceedings of WCCFL 22*, Somerville, MA: Cascadilla Press. 477–490.

Tesar, Bruce & Paul Smolensky (1998). Learnability in Optimality Theory. *Linguistic Inquiry* **29**. 229–268.

Tesar, Bruce & Paul Smolensky (2000). *Learnability in Optimality Theory*. Cambridge, MA: MIT Press.

Zuraw, Kie (2007). The role of phonetic knowledge in phonological patterning corpus and survey evidence from tagalog infixation. *Language* **83**. 277–316.