Learning allomorphs as a basis for morphophonemic learning

Bruce Hayes

Department of Linguistics

UCLA

Research question

- Suppose the right way to learn morphophonemics is to *learn allomorphs first*, before the phonology is known.
- How might this be done?
- And are there advantages to doing it this way?

The model abstractly characterized

Analytic task

- Devise a system that behaves like this:
- Given representative, glossed paradigms, it coindexes every segment with the morpheme that it belongs to.

Sample illustrative mini-language

- Person prefixes: /ni-/, /bi-/, /ri-/ alternating by Backness/ Rounding Harmony
- Stems: /kimen/, /kurat/, /petep/, /loran/
- Morpheme labels are given an arbitrary index:

nikimen	$\text{sing}_4\text{-1p}_1$	nukurat	$\text{swim}_5\text{-1p}_1$
bikimen	$\text{sing}_4\text{-2p}_2$	bukurat	$\text{swim}_5\text{-2p}_2$
rikimen	$\text{sing}_4\text{-3p}_3$	rukurat	$\text{swim}_5\text{-3p}_3$
nipetep	$\text{sit}_6\text{-1p}_1$	nuloran	$\text{think}_7\text{-1p}_1$
bipetep	$\text{sit}_6\text{-2p}_2$	buloran	$\text{think}_7\text{-2p}_2$
ripetep	$\text{sit}_6\text{-3p}_3$	ruloran	$\text{think}_7\text{-3p}_3$

Correct intended result

- Model should assign indices to segments thus, indicating the morpheme they belong to.

$n_1i_1k_4i_4m_4e_4n_4$	$sing_4\text{-}1p_1$	$n_1u_1k_5u_5r_5a_5t_5$	$swim_5\text{-}1p_1$
$b_2i_2k_4i_4m_4e_4n_4$	$sing_4\text{-}2p_2$	$b_2u_2k_5u_5r_5a_5t_5$	$swim_5\text{-}2p_2$
$r_3i_3k_4i_4m_4e_4n_4$	$sing_4\text{-}3p_3$	$r_3u_3k_5u_5r_5a_5t_5$	$swim_5\text{-}3p_3$
$n_1i_1p_6e_6t_6e_6p_6$	$sit_6\text{-}1p_1$	$n_1u_1l_7o_7r_7a_7n_7$	$think_7\text{-}1p_1$
$b_2i_2p_6e_6t_6e_6p_6$	$sit_6\text{-}2p_2$	$b_2u_2l_7o_7r_7a_7n_7$	$think_7\text{-}2p_2$
$r_3i_3p_6e_6t_6e_6p_6$	$sit_6\text{-}3p_3$	$r_3u_3l_7o_7r_7a_7n_7$	$think_7\text{-}3p_3$

- N.B. segments might not be adjacent (metathesis, etc.)

A note on the task

- This is related to, but not the same as, as the task of unsupervised learning of words and morphemes (e.g., Goldwater et al. 2009).
  - I'm modeling late, not early, acquisition.
  - Meant to be 100% accurate, rather than heuristic.
  - Meant to feed into wug-testing (ability to synthesize new forms).
  - Meant to be *fed by* unsupervised learning systems.

Coindexation as candidate selection

- GEN + EVAL, deployed with MaxEnt (Smolensky 1986, Goldwater and Johnson 2003)
- GEN = all possible coindexations of the segments
- EVAL uses various constraints of which the two main ones are as follows.

Faithfulness

- Adopt the well-motivated assumption (Kiparsky 1982, Steriade 2000, White 2016) that the final grammar should favor **minimization of allomorphy**.
- Let this be quantified thus: for each candidate, compile its “allomorph set” (like [ni-] ~ [nu-]) and assess violations of:
- **Faithfulness**: *penalize a candidate by the average dissimilarity of all allomorph pairs in the allomorph set.*

Calculating similarity

- I borrow the needed apparatus from:
- White 2016 (experiments → feature weights)
- Wilson and Obdeyn 2009 (feature similarity → segment similarity)
- Bailey and Hahn 2000 (similarity of best-aligned segments → string similarity)

Stem faithfulness

- Use separate, stronger version of Faithfulness for stems (McCarthy & Prince 1995, Beckman 1997)

Variegation

- Example of why this is needed; a segmentation that is *perfect* w.r.t. Faithfulness yet is wrong:

n-ikimen	‘sing 1p.’	n-ukurat	‘swim-1p.’
b-ikimen	‘sing 2p.’	b-ukurat	‘swim-2p.’
r-ikimen	‘sing 3p.’	r-ukurat	‘swim-3p.’
n-ipetep	‘sit 1p.’	n-uloran	‘think-1p.’
b-ipetep	‘sit 2p.’	b-uloran	‘think-2p.’
r-ipetep	‘sit 3p.’	r-uloran	‘think-3p.’

A constraint to enforce variegation in stems

- VARIEGATION

  Penalize candidate assignments in which stems begin or end with only a few different segments
- Measure employed: “worst case” count/total stems

Building a practical version of the model

20 Data Sets

- 10 are made-up languages, meant to pose some particular challenge to the system – such as the variegation language just given.
- 10 are problem sets from Kenstowicz and Kisseberth (1979): Bizcayan, Chamorro, Catalan, Polish, Lamba, Maori, Maltese, Lomongo, Okpe, Modern Hebrew.

Why problem sets?

- They include often-dramatic phonological alternations, intended to challenge the student.
- E.g. Okpe: [zũ] 'fan', but /ɛ-zũ-ɔ/ → [ezwõ] 'fan-inf.' with opaque harmony of [ATR], [round] and [nasal].
- Such dramatic alternations are the stuff of phonological theory, and perhaps will help us offer a useful empirical challenge to our friends in computer science working on similar problems (Cotterell et al. 2017).

Weighting the constraints

- They are fitted (Excel Solver) to pick linguist-selected winning candidates, from a large spreadsheet of plausible rivals from the 20 sample languages.
- Rival candidates come from my own thinking, failed earlier versions of the model, and losing candidates from the search beam.

Searching for winners

- Start by fixing the affiliations of non-alternating segments.
- Then for the others, do a hill-climbing beam search.
- Search moves: (1) single-segment edits on the data set, alternating with (2) edits of the allomorph sets.

Criterion of adequacy

- Allocate high probability (> .99) to the linguist-selected correct answer.

How is the model doing?

- All 10 made-up languages, plus 8 real languages: the outcome of the search is the correct answer.
- Okpe and Hebrew: search fails. But its best candidate is far less harmonic than the linguist's candidate, which I conjecture to be the true optimum.

The final step: full morphophonemic learning

Theme

- Full learning of the morphophonemic pattern becomes easier if you have the allomorph set in hand.

Begin by finding the alternations, with string-alignment

- String-aligning the discovered allomorph pairs, you can learn what the alternations are.
- English example:

i	t			[it]	'eat'
|	|				
i	ɾ	ɪ	ŋ	[iɾɪŋ]	'eating'

- Therefore: [t] ~ [ɾ] is an alternation.

Using the complete list of alternations

- You can use it to:
  - Rank the Faithfulness constraints (many can be immediately made undominated)
  - Construct a modest-sized yet comprehensive GEN function
- As a small-scale demo, let us use the allomorph sets to solve phonology problems.

Small-scale demo: full-UR and grammar learning, I

- First extract alternations from allomorph sets with alignment, as above.
- Then use this to make a defensible GEN: *apply all alternations to all allomorphs*.
- No winning candidate could ever be outside this set.
- No point in testing any other alternations.

Full-UR and grammar learning, II

- Code up a Universal Constraint Set for OT, following classical assumptions.
- (N.B. I and many people feel it would be better to learn the constraints instead.)
- See Hayes and Wilson (2008, *LI*) for learning of surface-true Markedness constraints.

Full-UR and grammar learning, III

- Try out all possible combinations of attested allomorphs as UR's, assuming the Single Surface Base Hypothesis (Albright 2002).
- The small GEN and small UR-set make it easy to check every choice.
- See Pater et al 2012, Jarosz 2015 for a more sophisticated methods of searching for UR's.

Full-UR and grammar learning, end result

- Correct UR's and ranking for Bizcayan, Chamorro, Catalan, Polish, Lamba, Maltese, Lomongo, and Okpe.
- This set is all the Kenstowicz/Kisseberth problems I tried that have purely-concatenative morphology (so, a simple GEN).

Planned next step: upgrade to wug-testing capacity

- Assertion: phonology problem set answers **grossly underestimate** the native speaker's knowledge.
- There are hundreds of **Islands of Reliability** (Albright 2002) that must be learned as well; local environments where particular outcomes are favored (work of Ernestus/Baayan, Albright/Hayes, Zuraw, Becker, Gouskova, etc.)
- So the next step is to augment the model to discover the Islands (see Albright and Hayes 2003 for one effort).
- This will make possible empirical work; i.e. modeling the behavior of humans in wug-testing.

Conclusions

- "Find the allomorphs first" seems to be a viable strategy for morphophonemic learning.
  - It appears to be **feasible**.
  - It likely has a smaller **search space** than alternative approaches (cf. Tesar 2014).
  - It is **needed for lexical allomorphy** anyway (Paster 2006).
  - It **eases** discovery of UR's and constraint rankings.

Thank you

- Email advice/feedback welcome: bhayes@humnet.ucla.edu

References I

- Albright, Adam. 2002a. Islands of reliability for regular morphology: Evidence from Italian. *Language* 78. 684-709.
- Albright, Adam. 2002b. The identification of bases in morphological paradigms. Ph.D. thesis, University of California, Los Angeles. www.mit.edu/albright/papers/AlbrightDiss.html.
- Albright, Adam. 2008b. Inflectional paradigms have bases too: evidence from Yiddish. In *The bases of inflectional identity*, eds. Asaf Bachrach and Andrew Nevins, 271–312. London: Oxford University Press.
- Albright, Adam. (2012). Base-driven leveling in Yiddish verb paradigms. *Natural Language and Linguistic Theory* 28: 475–537.
- Albright, Adam and Bruce Hayes (2003). Rules vs. analogy in English past tenses: A computational/experimental study. *Cognition* 90. 119-161
- Apoussidou, Diana (2007) *The learnability of metrical phonology*. Utrecht: LOT.
- Bailey, Todd M. and Ulrike Hahn. (2001). Determinants of wordlikeness: Phonotactics or lexical neighborhoods? *Journal of Memory and Language* 44, 568–591.
- Boersma, Paul (2001). Phonology-semantics interaction in OT, and its acquisition. *Papers in Experimental and Theoretical Linguistics* 6. Eds. Robert Kirchner, Wolf Wikeley & Joe Pater. Edmonton: University of Alberta. 24-35. [ROA 369, 1999].
- Della Pietra, Stephen, Vincent J. Della Pietra, and John D. Lafferty. (1997). Inducing features of random fields. *IEEE transactions on pattern analysis and machine intelligence* 19:380–393.
- Ernestus, Miriam and Harald Baayen (2003). Predicting the unpredictable: Interpreting neutralized segments in Dutch. *Language* 79, 5-38
- Goldwater, Sharon & Mark Johnson. (2003). Learning OT Constraint Rankings Using a Maximum Entropy Model. In Jennifer Spenader, Anders Eriksson & Östen Dahl (eds.), Proceedings of the Stockholm Workshop on Variation within Optimality Theory, 111–120. Stockholm: Stockholm University.
- Goldwater, Sharon, Thomas L. Griffiths and Mark Johnson. (2009) A Bayesian framework for word segmentation: Exploring the effects of context. *Cognition* 112:21–54.
- Hale, Kenneth L. (1973). Deep-surface canonical disparities in relation to analysis and change: an Australian example. In Sebeok, Thomas A. (ed.), *Current trends in linguistics 8: Linguistics in Oceania*, 401-458. The Hague/Paris: Mouton.
- Hayes, Bruce and Colin Wilson. (2008). A maximum entropy model of phonotactics and phonotactic learning. *Linguistic Inquiry* 39: 379-440.
- Hayes, Bruce, Kie Zuraw, Peter Siptár and Zsuzsa Londe. (2009). Natural and unnatural constraints in Hungarian vowel harmony. *Language* 85:822-863.

References II

- Jarosz, Gaja. (2006). Rich lexicons and restrictive grammars – maximum likelihood learning in Optimality Theory. Doctoral dissertation, Johns Hopkins University, Baltimore, Md.
- Jarosz, Gaja (2015) Expectation driven learning of phonology. Ms. Department of Linguistics, University of Massachusetts, Amherst.
- Kenstowicz, Michael and Charles Kisseberth (1979) *Generative Phonology: Description and Theory*. San Diego: Academic Press.
- McCarthy, John and Alan Prince (1995) Faithfulness and reduplicative identity. *Papers in Optimality Theory*, ed. by Jill N. Beckman, Laura Walsh, and Suzanne Urbanczyk, 249-384. (University of Massachusetts Occasional Papers in Linguistics 18.) Amherst, MA: Graduate Linguistics Students Association.
- Paster, Mary (2006). Phonological conditions on affixation. Ph.D. dissertation, UC Berkeley.
- Pater, Joe. 2009. Weighted Constraints in Generative Linguistics. *Cognitive Science* 33: 999-1035.
- Pater, Joe, Robert Staubs, Karen Jesney, and Brian Smith (2012) Learning probabilities over underlying representations. *Proceedings of the Twelfth Meeting of the Special Interest Group on Computational Morphology and Phonology* (SIGMORPHON2012), pages 62–71, Montréal, Canada, June 7, 2012.

References III

- Sankoff, David and Joseph Kruskal (1999) *Time Warps, String Edits, and Macromolecules: The Theory and Practice of Sequence Comparison*. Stanford, CA: Center for the Study of Language and Information.
- Smolensky, Paul (1986) Information processing in dynamical systems: Foundations of Harmony Theory. In James L. McClelland, David E. Rumelhart and the PDP Research Group. Parallel distributed processing: Explorations in the microstructure of cognition, Vol. 2: Psychological and biological models. Cambridge: MIT Press. 390-431.
- Steriade Donca (2000) Lexical conservatism and the notion base of affixation. Ms. University of California, Los Angeles
- Tesar, Bruce (2014) *Output-driven phonology*. Cambridge: Cambridge University Press.
- Tesar, Bruce, and Paul Smolensky. (1993). The learnability of Optimality Theory: An algorithm and some basic complexity results. Ms. Department of Computer Science and Institute of Cognitive Science, University of Colorado at Boulder. Rutgers Optimality Archive ROA-2, http://ruccs.rutgers.edu/roa.html.
- White, James (2014). Evidence for a learning bias against saltatory phonological alternations. *Cognition* 130. 96–115.
- Zuraw, Kie (2000). Patterned Exceptions in Phonology. Ph.D. dissertation, UCLA.
- Zuraw, Kie (2010). A model of lexical variation and the grammar with application to Tagalog nasal substitution. *Natural Language and Linguistic Theory* 28(2): 417-472.