Bruce Hayes
UCLA

Stanford talk
Apr. 2, 1998

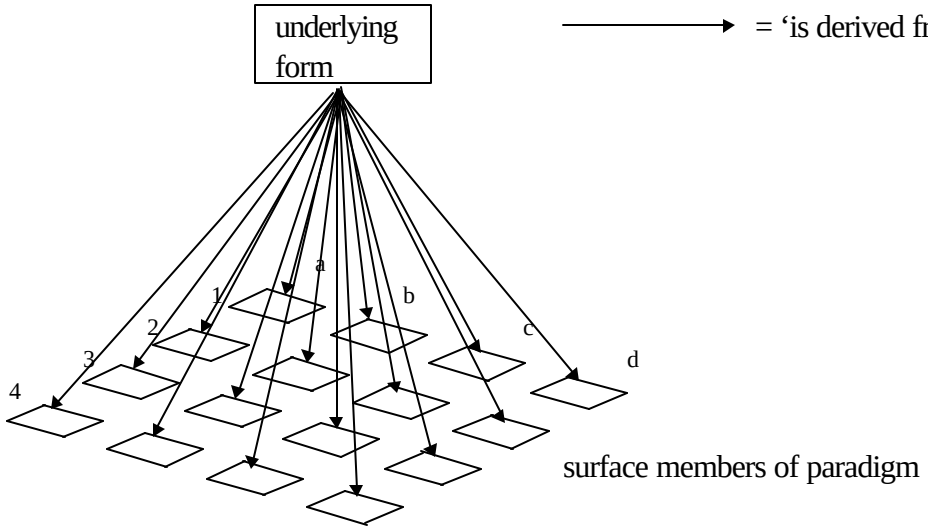
On the Richness of Paradigms, and the Insufficiency of Underlying Representations in Accounting for them

1. What Do Underlying Representations Do?

a restatement of ideas in Flemming (1995)

Basic Mechanism:

- All surface forms in the same paradigm are derived from the same underlying form.
- Surface resemblances across the paradigm are thus accounted for not directly, but as "theorems" of the derivational system.



Here, "derived from" can mean a wide variety of mechanisms (rules, GEN + candidate selection), depending on the theory.

2. Situating the Theory: The Wug Test (Berko 1958)

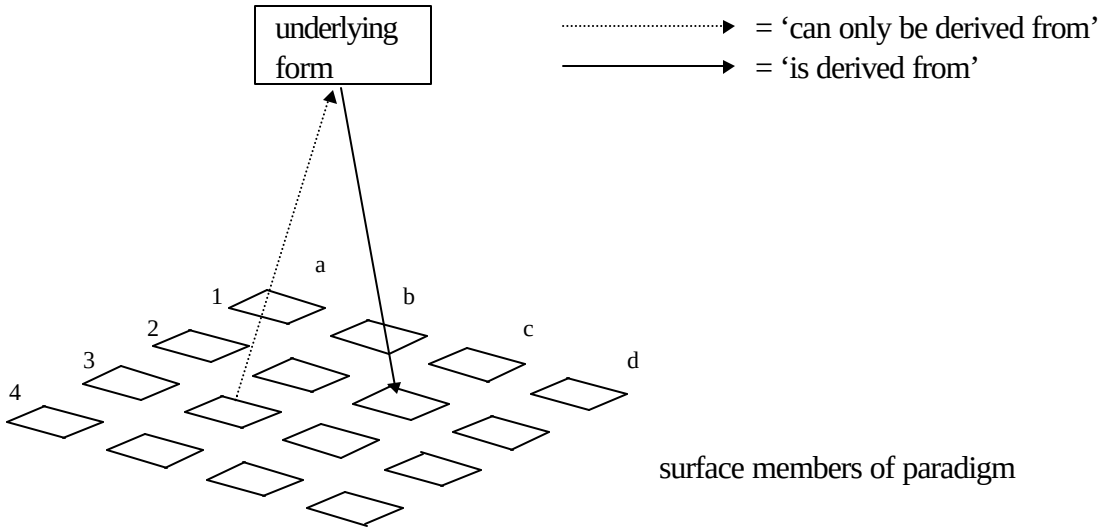
Given one member of a paradigm, provide another:

- "What is the plural of [wʌg]?"..."[wʌgz]"
- "If 'to the space alien' is [mulaninda], what is 'space alien?'" ... "[mulaːn]"

3. How a Speaker is Presumed to Pass the Wug Test (in theory of (1))

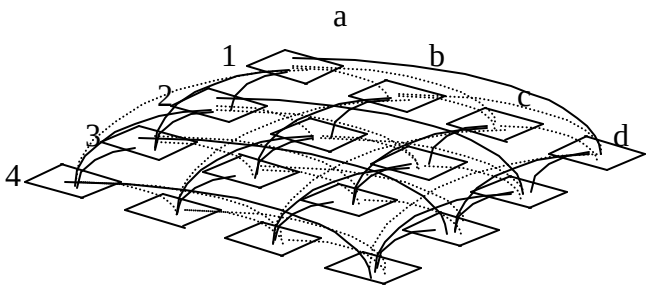
a) Use the morphological parser to detect the morphemes present, **undoing** the phonology.
b) Suitably add, subtract morphemes so as to **create the underlying representation** for the appropriate paradigm member.
c) Apply the phonology in the **forward** direction, and utter the output.

4. Wug Test: Project Paradigm Member [2C] from [3B]



5. An Alternative to Underlying Forms

Adopt a rich set of string mappings (incorporating phonology and morphology) that relate members of the paradigm in pairwise fashion. This tentatively proposed by Bochner (1993); see also the computational references cited by Sproat (1992, 215-216).



6. What Might Support A Rich Set of String Mappings?

A larger set of predictability relationships among the members of the paradigm than can be accounted from by deriving them from a single underlying form.

7. Main Idea of Talk

This indeed occurs, perhaps frequently.
Here, three cases: Korean, English, Yidiŋ

8. Before We Continue: A Non-Argument Against UR's

"UR's are abstract and unobservable and hence undesirable a priori."

The legitimate reason to abandon UR's is that they're not good enough.

KOREAN

9. Korean Neutralizations in Stem-Final Consonants

Before a vowel-initial suffix	Before pause or oral consonant
{[p], [p^h], [p']}	[p]
{[t], [t^h], [t'], [tʃ], [$tʃ^h$], [tʃ'], [s], [s']}	[t]
{[k], [k^h], [k']}	[k]

(There are also a few clusters that neutralize as well.)

10. Prediction of Traditional Theory (Derivation from Underlying Form)

Since the multiple consonants that appear before vowel endings all neutralize before pause, we should be massive differences in Wug-testing.

Mapping			**Wug-testing result**
pre-V allomorph	→	prepausal allomorph	unique answer
prepausal allomorph	→	pre-V allomorph	many possibilities

11. Some Tentative Results on Korean Wug-testing

(partly by me, partly by my undergraduate students)

- Wug-testing in the neutralizing direction (pre-V → prepausal) yields the expected single result.
- Wug-testing in the antineutralization direction (prepausal → pre-V), speakers frequently reject many of the logical possibilities.

12. A Korean Wug Test (Kang 1998)

"A [nut] is a tiny organism found in rotten food. Please fill in the blank:

[___ ɨl] məkimjən maʃi isaŋhata] 'If you eat *nut* it will taste strange'"

Consultant's[1] reply: [nusɨl]
Consultant's judgment on a 7 (perfect)-1 (bad) scale:

[nusɨl] **7** [nutɨl] **1** (various other consonants got intermediate values).
(Note: /t/ is allophonic [d] when between vowels.)

13. Reshaping of the Korean Lexicon

- My students repeatedly note that words with traditional stem-final consonants are acquiring innovating pronunciations ("We're supposed to say it this way, but really, everyone says this"). From Kang (1998):

Traditional	**Innovating**		
[ip] ~ [ip^{h}-ɨl]	[ip] ~ [ip-ɨl]	'leaf'	(Note: /p/ is allophonic [b] when between vowels.)
[nat] ~ [na$\widehat{tʃ}$-ɨl]	[nat] ~ [nas-ɨl]	'daytime'	(Note: /$\widehat{tʃ}$/ is allophonic [$\widehat{dʒ}$] when between vowels.)

- Martin's (1992) reference grammar, p. 101 fn.: "Many speakers treat the few nouns ending in a basic **t** as if they ended with an **s**. Even **tikut** 'the letter T' is pronounced with final **s** by most speakers when it is followed by [a vowel suffix]. But the Hangul spelling writes final **t** for this noun."

14. Conjecture

- There are principles of Korean phonology that permit speaker to project suffixed forms for stems that they have only heard in neutralized contexts.
- E.g. [t]-final isolation stems have as a default expectation that when affixed, they will appear with [s].
- As far as I can tell, the principles are **rational**: Korean is rich in /s/-stems, poor in /t/-stems, so [s] is a good guess.
- These principles are presently reshaping the lexicon, as more words are brought within their scope.
- See later on for what kind of theories might account for this phenomenon.

15. Underlying Representations can be Made to Work, for Korean

- **Constrain the underlying forms**: "Stem-final coronals must be (in the absence of compelling counterevidence) /s/."

[1] Kang's consultant. I have verified the judgments with my colleague Sun-Ah Jun.

- This solution fails, however, in other languages.

THE ENGLISH CASE

16. Attaching *-al* to [-ənt] Nouns: A Wug Test

"This, then, is the [ˈpɑkjədənt]. The [ˈpɑkjədənt] forms one of the central concepts of the religion of the Nefusites. Indeed, study of the [ˈpɑkjədənt] in the various religions of the world shows that **[___] al** concepts have arisen independently on many different historical occasions."

17. Characteristic Reply

[ˌpɑkjəˈdɛntal]

This is a rational reply, because virtually all [-ənt] stems show up with [ɛ] when *-al* is added. Forms below from a search of a smallish computer lexicon:

accident	[ˈæksədənt]	*accidental*	[ˌæksəˈdɛntəl]
coincident	[kəˈwɪnsədənt]	*coincidental*	[kəˌwɪnsəˈdɛntəl]
transcendent	[ˌtrænˈsɛndənt]	*transcendental*	[ˌtrænsˌɛnˈdɛntəl]
continent	[ˈkɑntənənt]	*continental*	[ˌkɑntənˈɛntəl]
orient	[ˈɔriənt]	*oriental*	[ˌɔriˈɛntəl]
parent	[ˈpærənt]	*parental*	[pəˈrɛntəl]
compartment	[kəmˈpɑrtmənt]	*compartmentalize*	[kəmˌpɑrtˈmɛntəlˌaɪz]
department	[dəˈpɑrtmənt]	*department*	[dəˈpɑrtˈmɛntəlˌaɪz]

+ 14 more *-ment* words

- Moreover, there are essentially **no** cases where any other vowel is restored from schwa, under *-al* affixation.
- The only possible exception is *consonantal*, but this is pronounced [ˌkɑnsəˈnɛntəl] by many linguists, despite orthography—surely this pronunciation is what needs to be explained.

18. A Second Wug Test

"[ˌpæləˈdɛntəl] concepts have also played a major role in polytheistic religions. Zoroaster, a major [ˌpæləˈdɛntəl] thinker, conceived the [ˈpæləd___nt] as being a manifestation of ..."

19. Answer

- Preferred responses seem to be [ˈpælə˛dɛnt], with no vowel alternation; and [ˈpælədənt], with schwa.
- Other vowels: impossible.

These judgments, too, are rational, reflecting the data pattern found in the lexicon.

president: [ˈprɛzədənt], [ˈprɛzə˛dɛnt]
orient: [ˈɔriənt], [ˈɔri˛ɛnt]

20. The Insufficiency of Underlying Forms

In normal analyses of English (dating from Chomsky and Halle 1968), schwa is derived by Vowel Reduction from stressless lax full vowels.

In this view, the [˛pæləˈdɛntəl] → [ˈpælədənt] mapping is trivial, reflecting the action of the stress rules and Vowel Reduction.[2]

But the [ˈpɑkjədənt] → [˛pɑkjəˈdɛntal] mapping is troublesome.

21. An Traditional Account: Constrain Underlying Forms

"[+syllabic] must be /ɛ/ in the environment / ___ nt]$_{word}$"

This is blatantly false: *stunt* [ˈstʌnt], *stint* [ˈstɪnt], *plant* [plænt], *gallivant* [ˈgælə˛vænt], *Dupont* [duˈpɑnt], etc.

More accurate would be:

"[+syllabic] must be /ɛ/ / ___ nt]$_{word}$ in adjectives and non-exceptional polysyllabic nouns"

= Precisely the contexts in which the English stress rules skip over final Vnt]

But even this doesn't really work:

[2] [˛pæləˈdɛntəl] → [ˈpælə˛dɛnt] also works smoothly, assuming some indeterminacy (Ross 1972) in the assignment of secondary stress to cluster-final words (secondary stress blocks reduction).

22. Last Wug Test

"The concept of the [ˈpɛljədənt] is likewise central to monotheism. Some world religions, though not adopting the [ˈpɛljədənt] as a doctrinal concept *per se*, nevertheless include parallel concepts with a definitely [ˌpɛljəˈd___ntɪk] flavor."

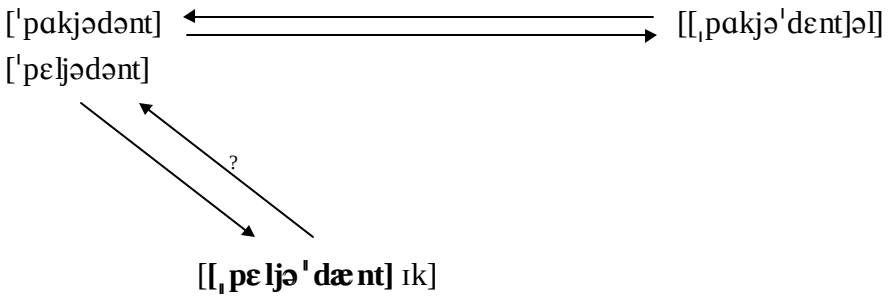
Here, there is no obvious preference for /ɛ/.
To the extent that [ˌpɛljəˈd**æ**ntɪk] is preferred, this seems to be based on parallelism with a single form, *pedantic*.

Note that there can be no rule ɛ → æ / ___ nt + ɪk], given the existence of *authentic* and *identical*.

Any theory that attributed the alternation to a *restriction on stems*, as above, could not get a different vowel restored in different contexts.

23. Upshot

The English vowel quality paradigm includes multiple relations of predictability among allomorphs, more copious than could be obtained by deriving all allomorphs from a single underlying form.



Theorizing should perhaps consider models in which these mappings are captured directly. See below.

24. A Common Thread in the English and Korean cases

- Phonological restructuring
- Restructuring makes it possible for the language learner/user to project novel affixed forms from neutralized isolation forms:
    - **Korean**: [nat] ~ [nat͡ʃ-ɨl] is being replaced by [nat] ~ [nas-ɨl], with the regular isolation → context mapping.

- **English**: [ˈkɑnsənə̲nt] ~ [ˌkɑnsəˈnæ̲ntəl] is being replaced by [ˈkɑnsənə̲nt] ~ [ˌkɑnsəˈnɛ̲ntəl], again with the regular isolation → context mapping.

YIDIŋ

25. Preview

Yidiŋ phonology appears to have been radically restructured, in a way the increases the projectibility of one paradigm member from another.

26. Yidiŋ (Australian, North Queensland)

All data, and virtually all generalizations and insights, from R. M. W. Dixon (1977) *A Grammar of Yidi*ŋ.

27. How the Conundrum Developed: Historical Yidiŋ

a) Apocope drops final vowels in absolutives (bare stems), but not in inflected forms
b) If this was like most sound changes, it probably was optional at first; this stage not too hard for learners.
c) But once it became obligatory, Yidiŋ became ripe for restructuring.
d) Yidiŋ learners of the restructuring generation modified the system, to increase projectibility of affixed forms.
e) Result: multiple directions of predictability, again richer than derivation from a single underlying form can handle.

28. Sound Change I: Penultimate Lengthening

In every word with an odd number of syllables, the penultimate vowel was lengthened.

29. Apocope

The final vowel of a word was deleted if:

- The resulting form would end in V + legal word-final consonant (/l,r,ɽ,j,m,n,ɲ,ŋ/)
- The resulting form would possess an even number of syllables.

30. The Metrical Component

- Virtually all metrical analyses of Yidiŋ (starting with Dixon's) assume disyllabic, left-justified feet, so that odd-even syllabicity predicates imply metrical structural descriptions.
- These feet also account for the alternating stress pattern.

31. Examples

a. With the accusative suffix [-na]:

[buɲa]	'woman-absolutive'
[buɲaː-n]	'woman-accusative' (hist. *[buɲa-na] → *[buɲaː-na] → [buɲaː-n])
[ŋunaŋgara]	'whale-absolutive'
[ŋunaŋgaraː-n]	'whale-accusative'

compare:

[gudaːga]	'dog-absolutive' (historically *[gudaga])
[gudaga-na]	'dog-accusative'

b. In stems:

[balbaːɽ]	~	[balbaɽ**a**-nda]	'crane-absolutive/dative'
[biguːn]	~	[bigun**u**-nda]	'shield-absolutive/dative'
[biniːr]	~	[binir**i**-nda]	'shell, money-absolutive/dative'

History: *balbaɽa > balbaːɽa > balbaːɽ

32. The Learnability Conundrum

According to Dixon, the absolutive form of a noun occurs in text more frequently than all other forms combined. Thus, for many stems, the following two questions confront the learner:

- For rarer words, which vowel is to be "restored" in the suffixed allomorph?
- How do you know whether to restore a vowel at all?

33. How New Yidiɲ Speakers Coped: The "Which Vowel" Problem

In Modern Yidiɲ, the following two ranked principles suffice to determine the "inserted" vowel:

a) If the final segment of the isolation stem is a **nasal**, the inserted vowel is **/u/** (14 cases, 0 exceptions). Example: [baɽːn] 'tree used for handles-absolutive' ~ [baɽin**u**-nda] 'dative'

[baɽiɲ(**u**)]	'tree used for handles'	[gulan(**u**)]	'walnut tree'
[dʲalam(**u**)]	'fresh, young'	[guɽban(**u**)]	'crow'
[dʲurin(**u**)]	'leech'	[malan(**u**)]	'right hand'
[gaban(**u**)]	'rain'	[muɽin(**u**)]	'ashes'
[gambin(**u**)]	'top-knot pigeon'	[wangam(**u**)]	'kidney'
[gindan(**u**)]	'moon'	[waŋgam(**u**)]	'overhanging cliff'

[gugi**ɲ**(**u**)] 'flying fox' [wuga**m**(**u**)] 'firefly'

b) Otherwise, the inserted vowel is a **copy of the rightmost vowel** of the isolation stem: thus [waji:l] ~ [waji**l**-nda] 'red bream-absolutive ~ dative'.

[waj**i**l(**i**)] 'red bream'
[win**a**ɽ(**a**)] 'foot'
[wiɽ**u**l(**u**)] 'a shell fish'

34. A New Aspect of the Data

- Dixon, lacking Microsoft Excel, noticed (33b) but not (33a).
- Once (33a) is added to (33b), the ability of speakers to project the added vowel becomes very strong, suggesting a quite different analysis.

35. Historical Change: Forms were Shifted Toward Obedience to (33)[3]

Yidiɲ		**Cognate**
[gaw**u**l(**a**)/(**u**)]	'blue gum tree'	[gawul**a**] (Dya:bugay)
[mag**u**l(**a**)/(**u**)]	'a root food'	[magul**a**] (Dya:bugay)
[ɲag**i**l(**a**)/(**i**)]	'warm'	[ɲagil**a**] (Gunggay), [ɲigal**a**] Mamu Dyirbal)
[jag**uɲ**(**u**)]	'echidna'	[juguɲ**a**ŋ] (Gunggay)
[muɽi**n**(**u**)]	'ashes'	[murin**i**] (Dya:bugay)
[band^{j}**a**ɽ(**a**)]	'madness in head'	[band^{j}**a**ɽ] (Dyirbal)
[d^{j}uŋg**um**(**u**)]	'worm'	[d^{j}uŋgum] (Dya:bugay)

36. The Residual Exceptions (8 total)

[d^{j}amb**u**l(**a**)]	'two'	[gub**um**(**a**)]	'black pine'
[gang**u**l(**a**)]	'grey wallaby'	[gul**a**ɽ(**i**)]	'big-leaved fig tree'
[gambi**ɽ**(**a**)]	'tablelands'	[guŋg**a**ɽ(**i**)]	'north'
[gaɲd^{j}**i**l(**a**)]	'crab'	[waŋ**a**ɽ(**i**)]	'pre-pubescent boy'

I assume these are now lexically-listed allomorphs.

37. Where did the Patterns of (33) Come From?

They appear to have been **statistical accidents of the existing lexicon**—this can be determined by inspecting trisyllabic stems that were never phonologically eligible for Apocope in the first place; in these, the patterns are **weakly present as tendencies**.

→ Yidiɲ learners seized upon the "best guess" strategy for restoring the lost vowels.

[3] There are a couple of glitches, where I think it can be shown that the pressure toward change is from a minor sub-generalization (compare English *dive ~ dove*). See full paper for details.

38. The "Whether to Insert" Problem

Background: Pre-Yidiŋ had very few phonemic (non-alternating) long vowels. A near-minimal pair:

'person'	'mopoke owl'
[bam**a**]	[durg**u:**]

Today, just 14 morphemes like [durgu:]; none with [...V:C]

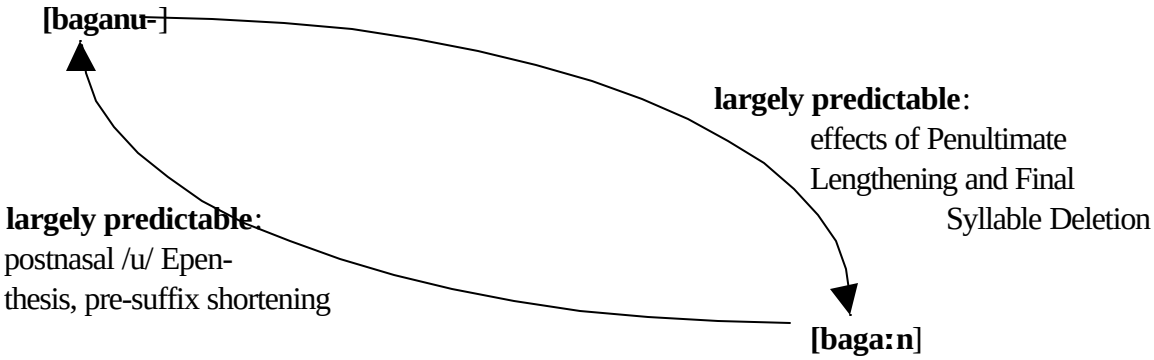
→ It is certain that any form ending in V:C will get an additional vowel under suffixation.

E.g. if you know absolutive [gawu:l], you **know** the inflected form will be [gawulV-CV]

If there **ever were** forms that were like "[bawu:l] ~ [bawu:l-CV]", they must have been shifted into the now-regular pattern, following the majority.

39. Yidiŋ Paradigms are Rich Paradigms

- The old analysis (Penultimate Lengthening, Apocope) **still works** (though with a certain number of lexical exceptions).
- The new system of "epenthesis" is **overlaid** on the old system.
- Thus Yidiŋ, too, is a case where there is **more predictability** within the paradigmatic pattern than you can get from deriving the forms from a single UR.



40. Could we account for paradigm richness with constraints on underlying forms?

We try:

a. In a trisyllabic root, vowels must be [u] / [+nasal] ___]
b. Else they must be a copy of the closest vowel on the left.

This would keep the old system. By limiting the set of underlying forms, we limit the possible range of alternation:

*"[babaːn] ~ [babani-ŋgu]" because UR would have to be illegal /babani/.

This fails: the constraints are not true of

a. Trisyllabic roots that are lexical exceptions to Apocope:

[galga**li**] 'curlew'
[gaɽa**na**] 'black cockatoo'

b. Trisyllabic roots that don't undergo Apocope because the last consonant is an obstruent, e.g. [bind**u**b**a**] 'crayfish-absolutive':

Local conclusion: the principles of vowel restoration govern **patterns of alternation**, not the set of underlying forms.

41. Further Evidence Against the Constraints-On-Underlying-Forms Strategy

The "restored vowels" given above are only the **primary**, most-frequent strategy.
There are other less frequent options, more or less in free variation.

a) Inserted vowel is copy of closest **suffix** vowel:

[gambiːn]	'top-knot pigeon-absolutive'
[gambin**u**-ŋg**u**]	ergative
[gambin**a**-l**a**]	locative
[gambin**i**-j**i**]	comitative

b) Inserted vowel is a **schwa** (not otherwise present in Yidiɲ!)

[gubuːm]	'black pine-absolutive'
[gubum**ə**-ni]	genitive

- These cases strike me as fatal to the constraints-on-underlying-forms approach. The variation is systematic, meaning it should be attributed to the rule system, not to the lexicon.
- This confirms approach of multiple mappings within the paradigm: these cases can be accounted for as simple free variation in one of the mapping directions.

CONCLUSIONS

42. The Theme in All Three Cases

- A neutralizing mapping (Korean: Consonant neutralization, English: Vowel Reduction, Yidiŋ: Apocope) applies in isolation forms, but not affixed forms.
- This engenders **reanalysis**: new patterns arise that permit the affixed forms to be projected from the isolation forms (Korean t → s / ___ + V, English ə → ɛ / ___ ntəl, Yidiŋ ∅ → u / V: [+nasal] ___ X]).
- I suspect that this phenomenon is quite common[4].
- It's common, but it's also **understudied**: the standard model that derives forms from underlying representations mistakenly tells the analyst "you're done!" when in fact there are more regularities that should be extracted. A better model might lead to better and more complete elicitation work.

43. There is something sensible about multiple mappings.

- It is useful to be able to say inflectional and derivational forms you have never heard before.
- Ways of deriving such novel forms, even if valid only statistically, let you talk more effectively than you otherwise could.
- This is all the more crucial for beginning learners, who must Wug-test their way through daily life.

44. Local Conclusions

- It is not unreasonable to think that language learners develop grammars that extract the full set of predictability relations from paradigms, even where the conventional device of derivation from a single underlying form can't do this.
- It is not unreasonable to consider abandoning underlying forms, and develop theories of phonology that aim from the beginning at extracting the full set of predictability relations.

45. Some Speculation on a Theory that Extracts Full Predictability Relations

- Speakers adopt **mapping constraints** that enable them to predict one member of a paradigm from another. These are string mappings that formalize the arrows of (5). Example: [Xst] → [Xən] (German 2nd sg. to 1 plur. verbs, *du* [zɪŋ**st**] 'you sing', *wir* [zɪŋ**ən**] 'we sing').
- Speakers do this for **all** mappings, unless the paradigm is large, in which case they map forms that differ only in one or two inflectional features.
- **Phonology helps**: if we can derive German /bliːb/ 'I stayed' from [bliːbən] 'we stayed' by [Xən] → [X], then Final Devoicing will get us the rest of the way, to [bliːp].

[4] For more, see Hayes (1995) on Turkish Final Devoicing, Kaye and Nykiel (1980) for Algonquian and Polish, Blevins (1997) on English and Gilbertese.

- Speakers memorize **enough** inflected forms in each paradigm to be able to project the others; specifically, in cases of phonological neutralization they often need to memorize at least one allomorph that escapes the neutralization.
- Exactly **what forms are memorized** is an accident of a speaker's personal history (compare Smolensky 1993 on underspecification: there, too, it doesn't really matter, within limits, what you memorize.)

46. Why Such a Theory is Likely to be Frightening to Work On

- It will most likely involve some redundancy.
- Hence the intuitive guides of elegance and economy that have hitherto helped us with analysis might not help here.

47. A Possible Way to Explore Such Theories

Develop algorithms for phonological learning. Criteria for success:

- The completed grammar that the algorithms learn should give Wug-test answers identical to those of adult native speakers of the target language.
- In the course of acquisition, the algorithm should make the same errors that children learning the target language make.

48. Why this might count as redemption

- While we might lose some elegance, the loss might be more than compensated for by a novel ability actually to **model the speaker** (rather than just the data pattern), and to explain phonological acquisition.
- Algorithmic learners tend to succeed or fail pretty clearly; hence questions of elegance arise less often in evaluating them—**anything** that actually works on hard cases would be very impressive, and at least somewhat persuasive.

49. A Debt to Pay

Algorithmically learned analyses of Korean, English, Yidiɲ ...

50. Advertisement

For tiny baby steps in this direction, see Albright and Hayes, "An Automated Learner for Phonology and Morphology," handout for Hayes presentation at this weekend's Germanic Linguistics Roundtable.

References

Berko, Jean (1958) "The child's acquisition of English morphology," *Word* 14, 150-177.

Blevins, Juliette (1997) "Rules in Optimality Theory: Two Case Studies," in Ignacio Roca, ed., *Derivations and Constraints in Phonology*, Clarendon Press, Oxford.

Bochner, Harry (1993) *Simplicity in Generative Morphology*, Mouton de Gruyter, Berlin.

Chomsky, Noam and Morris Halle (1968) *The Sound Pattern of English*, Harper and Row, New York.

Dixon, Robert M.W. (1977), *A Grammar of Yidiŋ*, Cambridge University Press, Cambridge.

Kang, Sarah (1998) "Linguistics 120A Term Paper: Aspects of Korean Phonology," ms., UCLA.

Flemming, Edward (1995) *Auditory Representations in Phonology*, UCLA dissertation.

Hayes, Bruce (1995) "On What to Teach the Undergraduates: Some Changing Orthodoxies in Phonogical Theory," *Linguistics in the Morning Calm 3*, Hanshin, Seoul.

Hayes, Bruce (1997) "Anticorrespondence in Yidiŋ," available at Web site listed below. Shorter version to be published in the Proceedings of the 1996 Tilburg Derivational Residue Workshop, ed. by Ben Hermans and Marc van Oostendorp.

Kaye, Jonathan and Barbara Nykiel (1979) "Loan Words and Abstract Phonotactic Constraints," *Canadian Journal of Linguistics* 24, 71-93.

Martin, Samuel E. (1992) *A Reference Grammar of Korean*, Charles E. Tuttle Company, Rutland, VT.

Ross, John R. (1972) "A Reanalysis of English Word Stress," in Michael Brame, ed., *Contibutions to Generative Phonology*, University of Texas Press, Austin.

Smolensky, Paul (1993) "Harmony, Markedness, and Phonological Activity," Rutgers Optimality Archive 87, http://ruccs.rutgers.edu/ROA/search.html.

Sproat, Richard (1992) *Morphology and Computation*, MIT Press.

Department of Linguistics
Box 154302
UCLA
Los Angeles, CA 90095-1543

bhayes@humnet.ucla.edu

www.humnet.ucla.edu/humnet/linguistics/people/hayes/hayes.htm