

Class 1.2, 7/2/03: Phonotactics I

1. Assignments

First-week readings
Study guide questions due Monday 7/7/03 in class
Optional OTSoft assignment handed out today

MODELING THE LEARNING OF PHONOTACTICS

2. Why It's Hard

- No negative evidence: you're "told" what's there, but learn what is not there...
- Need to generalize: keeping a list of all forms you've heard obviously won't suffice

[ɹɪs] [mɪm] [sɛg] [tɛt] [jæn] [ɹæɪ] [gɛn] [fɛk] [sɛʃ]
sound very good indeed, I think.

- Gradience: assuming we aim high and attempt to model gradient judgments of well-formedness, we need mechanisms of gradient learning
- The need for robustness: give less weight to exotica in the learning data:
 - [dweɪn], Harris's Spanish example *aunque* [ˈaũNke].

3. Two Strategies for Learning What is *not* there

- Make a list of all the things that could be "wrong with" a phonological representation. Let the child know this list and check it against what she has heard: OT (oversimplified)
- Find a way to calculate what you would *expect* to encounter, and check for gaps: the "observed/expected" method

LEARNING PHONOTACTICS IN OPTIMALITY THEORY

4. References

- Hayes reading
- Prince, Alan & Bruce Tesar (1999). "Learning phonotactic distributions." Rutgers Optimality Archive 353, <http://rucss.rutgers.edu/roa.html>. To appear in René Kager and Wim Zonneveld, eds., *Fixing Priorities: Constraints in Phonological Acquisition*, to be published by Cambridge University Press.

A possible grammar for this:

{ *VOICED SIBILANT >> AGREE(voice) >> IDENT(voice)

/sb/	* VOICED SIBILANT	AGREE(voice)	IDENT(VOICE)
☞ [sp]			*
*[sb]		*!	
*[zb]	*!		*

8. This is Probably Not the Way Children Do It

Reasons:

- Quite a bit of low-Faithfulness ranking, perhaps most, is *not* diagnosable by alternations.

E.g. *[Δ in content words

The alternations that would be needed to learn this would be:

/ ^l ΔIp↔ ₁ teIt/	
[^l θIp↔ ₁ teIt]	‘to sharpen a quill pen’
[ri ₁ ΔIp↔ ₁ teIt]	‘to sharpen a quill pen again’
/ ^l θIp↔ ₁ neIt/	
[^l θIp↔ ₁ neIt]	‘to dry with blotting paper’
[ri ₁ θIp↔ ₁ neIt]	‘to dry with blotting paper again’

But there are no stems at all like /^lΔIp↔₁teIt/.

- Phonotactic learning is precocious (e.g. 8 months), and in all probability occurs before morphology is learned; see later on in the course (indeed, it appears to occur without knowledge of the meanings of most of the words).

9. “Size” of Phonotactics Depends on how High Faithfulness is Ranked (Smolensky 1996)

- All Faith on top: all of GEN, or all that relates to some Faithfulness constraint, is sayable.
- All Faith on bottom: only the least marked candidate of GEN (depending on how Markedness constraints are ranked) is sayable.
- The huge number of intermediate rankings each define a phonotactics of an intermediate degree of rigor.

10. Defining the Goal

Rank Faithfulness “as low as possible”, compatible with the input data.
...or more precisely, as low as people do.

11. Constraint Demotion (Tesar and Smolensky 1993 et seq.)

- Tesar, Bruce and Paul Smolensky (1993) “The learning of Optimality Theory: An algorithm and some basic complexity results,” ROA-52, Rutgers Optimality Archive, <http://ruccs.rutgers.edu/roa.html>.
- Tesar, Bruce (1995) *Computational Optimality Theory*, U. Colorado dissertatio, ROA-90.
- Tesar, Bruce and Paul Smolensky (1996) “The learnability of Optimality Theory,” (much longer, with ideas about accessing underlying forms) ROA-156.
- Tesar, Bruce and Paul Smolensky (1998) “Learnability in Optimality Theory,” *Linguistic Inquiry* 29:229-268.
- Tesar, Bruce and Paul Smolensky (2000) *Learnability in Optimality Theory*. Cambridge: MIT Press.

Given

- input-output pairs
- appropriate illegal rival candidates
- a constraint inventory
- violation counts for all winners and rivals

This algorithm will find a ranking, if there is one, that derives only winners.

12. “Prefers a Loser”

- A constraint **prefers a loser** if at least one losing candidate violates it fewer times than the winner does.

13. Strata

Assign constraints to strata 1, 2, 3, ... n, such that if Constraint A is in a higher stratum than Constraint B, then A outranks B.

Rankings within strata can be anything you like; empirical results are the same.²

14. Constraint Demotion (simplest version)

Find all constraints that never prefer a loser.

Put them in Stratum 1.

Throw away all losing candidates that are explained by the constraints of Stratum 1 (i.e. which are dispreferred to a winner by one of the Stratum 1 constraints).

Among the remaining data, find all constraints that never prefer a loser.

Put them in the Stratum 2.

² This is an approximation; in more interesting grammars free ranking within strata can give rise to free variation.

Throw away all losing candidates that are explained by the constraints of Stratum 2.

Among the remaining data, find all constraints that never prefer a loser.

Put them in Stratum 3.

Throw away all losing candidates that are explained by the constraints of Stratum 3.

etc.

If at the end, all loser are explained, then only winners will win, and you have a working grammar.

15. Crashing

I. You find that every constraint prefers some loser.

II. You've exhausted all the constraints, and some losers are still tied with their winners (have the same violation pattern).

16. Strategy for Phonotactic Learning

- Let underlying representations be surface representations, by fiat (Prince/Tesar/Albro)—only alternations would tell us otherwise, and without morphology we have no alternations.
- Let the maximally low ranking of Faithfulness be the result of the ranking algorithm, not evidence from alternations.

17. Normal Constraint Demotion Can't Be Used here

- There are no faithfulness violations in winners, so Faithfulness constraints never prefer the loser, and thus can't be demoted.
- Resulting grammar allows *anything*.

BIASING CONSTRAINT DEMOTION TO PLACE FAITHFULNESS LOW

18. Basic Approach

Place as many (sensible) barriers as we can to placing a Faithfulness constraint in the stratum under construction.

When we choose among non-loser-preferring constraints to place in a new stratum, follow these selection criteria (blend of two approaches):

19. Favor Markedness (Hayes forthcoming, Prince/Tesar forthcoming)

- If it is possible to install a Markedness constraint in the newly created top stratum, do so, and don't install any Faithfulness constraint.

20. Favor Activeness (Hayes, Prince/Tesar)

- If, among Faithfulness constraints, you have some that prefer winners, and some that don't, don't install the ones that don't.

21. Favor Specificity (Hayes)

- If two winner-preferring Faithfulness constraints fall into a specific-general relation, place only the more specific one in the current stratum.

22. Favor Autonomy (Hayes)

- Of the constraints that satisfy all higher-priority ranking principles, install on the current stratum the constraint or constraints who most rule out losers without assistance from other constraints.

23. Favor Markedness Liberation (Prince/Tesar)

- Pick Faithfulness constraints to rank that maximally “free up” Markedness constraints to be ranked later on, when the algorithm is run in speculative forward mode.

24. Performance of These Models

Both models (if Prince/Tesar is amplified to include Favor Specificity) are good enough to handle toy examples; see Appendix to Hayes paper for a bake-off.

But I think both models have enough drawbacks that they count only as conceptual studies:

25. The Gradience Problem

- Stochastic algorithms, like Boersma's Gradual Learning Algorithm (Boersma and Hayes, *Linguistic Inquiry* 2001) can capture gradience in input/output relations.
- E.g. for Tagalog, every [f] is somewhat dubious, and is replaced by [p] in unguarded speech: [filipino/pilipino].

At least at a descriptive level, we have:

*LABIAL FRICATIVE <> IDENT(continuant)

at some level of probability defined in the theory of stochastic grammar.

- But no one has yet successfully adapted the Gradual Learning Algorithm or a similar gradient algorithm to do gradient phonotactics. See the optional exercise if you'd like to give it a try!

26. Reliance on Universal Markedness Constraints

- This is famously controversial: how much explicit information about phonological well-formedness is in the genome and how it could have gotten there...
- It also implies that if a lexicon had an “accidental” phonotactic it could not be learned.

27. Are There Accidental Phonotactics in English?

Candidates:

- 1)a. [ʊk] is ok (*book, look, rook, hook, took, cook, shook, nook*) but *[ʊp] is not
- b. [ɔk] is ok (*talk, walk, Montauk, Salk, Sauk, balk, chalk, caulk, gawk, auk*), *[ɔp] is not.
- c. [u:p] after a non-coronal is ok (*poop, coop, goop*) but [u:k] is not (expressive words only; *spook, gook*)³

These seem related to difficulty of maintaining a /p/-/k/ contrast after the low F2 of back rounded vowels. But choice of *which* consonant to avoid is perhaps arbitrary.

- 2) limitation of /ʒ/ to ambisyllabic and occasional word-final position. Neither [ʃ] nor [v] is thus limited. /ʒ/ is the result of now-opaque palatalization processes (*vision*) and a handful of borrowings (*garage, rouge*)

If there are accidental phonotactics, then ability to *construct constraints*—or engage in some other form of distributional learning—would be crucial.

28. The Markedness Subset Problem

The OT-based approach relies on a hypothesis about markedness in OT that is little commented on.

Suppose segment B has a *superset* of the Markedness violations of A.
Then

- B can occur in the output inventory only as a result of high-ranked Faithfulness.
- A will occur in the output as well.

Flemming (forthcoming)⁴ points out some stop inventories that are interesting from this point of view.

San Juan Colorado Mixtec stops

³ The post-noncoronal restriction is needed because there are [u] that come historically from [ju], via the drop of /j/ after coronals: *duke, nuke, fluke*.

⁴ To appear in Hayes, Steriade and Kirchner (in press) *PhoneticallyBased Phonology*, Cambridge U. Press.

π	τ	τθ	κ
^m ɓ	<δ	<δθ	

Nyangi Stops

p	t	c	k
ɓ	ɗ	·	ɠ

where prenasalized and implosive stops are commonly thought to be more marked than plain voiced.

RAW INDUCTIVISM: DIPHONES AND TRIPHONES

29. Core Idea

Perhaps all of the above problems will disappear if we abandon all this theoretical sophistication, roll up our sleeves, and learn the language...

30. The Simplest Strategy

- Chop the learning corpus into **phones** (segments), **diphones** (two-segment sequences) and **triphones** (three-segment sequences).
- The knowledge of phones provides the schema, namely all possible 2- and 3- segment concatenations of phones.
- Idea: any word that contains a non-existent diphone/triphone will be ill-formed. Any word resolvable entirely into “good” di/triphones will be well-formed.
- Intermediate frequency di/triphones may correspond to intermediate well-formedness.

31. Reference

I believe the idea is quite old, but for a sophisticated current discussion, see

- Pierrehumbert, Janet (in press) “Probabilistic Phonology: Discrimination and Robustness,” in R. Bod, J. Hay and S. Jannedy (eds.), 2002. *Probability Theory in Linguistics*. MIT Press (to appear). [<http://www.ling.nwu.edu/~jbp/publications.html>]

32. Giving It a Try

My corpus: an electronic desk dictionary. 41697 distinct entries, perhaps half the vocabulary of an educated English speaker.

This is smaller than one could do with official sources (e.g. CELEX), but presumably much larger than an infant could access.

Chopping:

- Remove duplicate entries from the word list.
- Prefix every word with the symbol “[“ and suffix with “]”, to be able to count initial and final sequences; i.e. [tri] yields [tr, tri, and ri] as triphones.
- Collect all distinct phones, diphones, triphones; along with their frequencies (i.e., frequency in a type-ized lexical corpus).
- Also compute and list the logically possible diphones and triphones that have a zero corpus frequency.
- Determine *proportional phone frequencies* = total number of a phone/total number of phones
- Determine *expected frequency value* of every diphone and triphone, defined as the product of the proportional phone frequencies of its members.⁵ For the triphone “tri”:

$$\text{Expected(tri)} = \text{Freq(t)} * \text{Freq(r)} * \text{Freq(i)} = 0.0524 * 0.0621 * 0.0293 = .0000953$$

33. Basics

Number of distinct phones:	42
Number of distinct diphones:	1681
Number of distinct triphones:	67240

The latter is slightly more than Microsoft Excel can examine; I simply ignored the small residue.

34. The Five Most Common Diphones

Diphone	Frequency	Fraction of total
ən	7835	0.023
ər	7348	0.022
əl	6235	0.018
t]	5514	0.016
n]	5173	0.015

35. Digression: This is a fine way to discover constraints!

These were new to me, and seem quite valid.

- No lax vowels before glides (*[EwA])
- No lax vowels before [h] (*[EhA])⁶

⁵ A more sophisticated approach would be to model an expected triphone frequency as the (best?/worst?) combination of diphone + phone.

⁶ This is perhaps “deducible”: h in onsets only; lax vowels always require at least an “ambi-coda”.

- c. No “voice-abstracted” geminates (*[fv])
- d. No “sequential imitations” of affricates: *[tʃ, dʒ]. *Betsy*, **Betsy*
- e. No sequential imitations of diphthongs: *[AjV], *[bōūw↔]. Cf. *lawyer*, *ayatollah*, with conversion to diphthongs, vs. *hallelujah*.

36. Problem I: Rarest Diphones Sound Pretty Good

Hapax legomena (= frequency 1) are mostly mistranscriptions and compounds (e.g. [dk] *woodcat* [dʒʃ] *judgeship*, [mθ] *something*).

Authentic hapaxes:

[u□]	in <i>fluor</i>
[ðA]	in <i>fathometer</i>
[ʒm]	in <i>Cashmere</i>
[aök]	in <i>Araucanian</i>
[zaü]	in <i>resound</i>
[dʒt]	in <i>vegetable</i>

with 2 instances:

[aIaI]	<i>myiasis</i>
[gʒ]	<i>luxuriant</i>
[tv]	<i>Latvia(n)</i>
[aüdʒ]	<i>gouge</i>
[eIE]	<i>Azrael</i>

Crucial point: there’s no lowest threshold that defines what is legal. I.e. it would be convenient if every legal diphone had at least 20 tokens!

This can be taken further:

37. Missing Diphones in Corpus For which I Could (Maybe) Think Of a Word

[aui]	<i>Maui</i>
[aug]	<i>Haugen</i>
[□I↔]	<i>Hoya</i>
[□Δ]	<i>Cauthen</i>
[tʃ□I]	<i>Choi</i>
[Nv]	<i>Ingvaeonic</i>
[w□I]	<i>quoits</i>

38. Example of Phonotactic Principle Missed by Diphones: Diphones that are Legal only Intervocally

or intersonorantically; something along these lines

[dl] *Adler, Bowdler, bedlam, Adlai, maudlin, medley, Ridley*, but never *[dlAp] or [Eln.dlAp]⁷ or **[Edl.bAp] or *[Adl]

[mb] *ember* and many others, but not *[mba], *[Emb.də], *[Imb], *[Iz.mbə]

Intuition: licensing contexts for sounds often involve more than one neighboring segment.

39. Result

We have to move up to the triphones!

40. Most Common Triphones

ən]	3431	0.012	
əl]	3045	0.010	
ər]	2755	0.009	a phoneme, in many analyses
əs]	2538	0.009	
tər	1670	0.006	
ʃən	1550	0.005	from <i>-ation</i> ?

41. Encouragingly Missing Triphones

- Sure enough, [dl, dl] are missing triphones, and (by hand checking) we find that that *lndl* is a missing tetraphone.
- [mb, mbd, mb], [zmb] are missing triphones.

42. But the Triphone Inventory Suffers from Countless Accidental Gaps

Just a few of them, some with examples I thought up, from outside the corpus:

[zEv]	(byname of Los Angeles politician)
[aɪdʒæ]	(exists for linguists: <i>IJAL</i>)
[wEm]	(<i>Wembley</i>)
[wɪdʒ]	(<i>widget</i>)
[veɪz]	(<i>vase</i> , dictionary gave [veɪs])

⁷ Medial /ndl/ occurs in *Handlin*, not in the corpus.

43. Observed/Expected Works Better

Frisch, Broe and Pierrehumbert (1997, ROA) for an important application of this idea, also Frisch, S. A., Large, N. R., Zawaydeh, B., & Pisoni, D. B. (in press). Emergent phonotactic generalizations in English and Arabic. In J. Bybee & P. Hopper (eds.), *Frequency effects and emergent grammar*. Amsterdam: John Benjamins.

Here, “expected” is the multiplied probabilities of the included phonemes

Sort: find zero-frequency diphones

Among these, find the 30 with the highest expected value. In other words, the “most convincing” zeros.

æə, əæ, εə, əε, əo, əɔ, æi, εi, əu, w], æi, h], ie, ei, iy, oi, ææ, tz, æε, εæ, tʃ, ɪɔ, [ŋ,

sz, wr, ʃs, εε, aæ, εa, aε

These look fairly legit, and it is not hard to find general, valid constraints of English that they violate.

44. Observed/Expected Might Not Help with /dl/

- There are not very many *medial* /dl/, especially if you exclude the compounds.
- So, the assertion “no /dl/ are initial” is singularly unimpressive—given the number of items, it could easily be due to accident.
- It all seems to depend on how we calculate the expected frequencies.

45. Tentative Conclusion

- Diphones aren’t big enough. There are well-attested diphones whose well-formedness depends strongly on their context.
- Triphones are too big. The huge range of possibilities they imply can’t be reliably searched with the size of lexicon that we have.
- Observed/expected helps, but I doubt it will help enough.
- It’s possible that an adequate theory of phonotactics will require some use of phonological theory (!).

46. A Syllable-Based Approach: Pierrehumbert and Coleman (1997= readings)

Their approach is similar to the diphone/triphone approach, but with more principled units.

- a) Find all possible onsets by considering all possible word-initial clusters.
- b) Parse all learning words (disyllables only) into onsets and rhymes, following the (dubious) Maximal Onset principle (e.g. *Atwater* is $\Theta.tw\alpha.t\alpha$).

- c) Calculate probabilities of onsets and rhymes for initial/final stressed/unstressed syllables, giving probability for word as a whole.

47. Possible Problems for Coleman and Pierrehumbert

- Dependencies that don't follow from CP's approach are common crosslinguistically:
 - between onset and rhyme,
 - between rhyme and following onset
 - between adjacent rhymes (null onset)

are fairly common crosslinguistically.

Examples:

- In many retroflexion languages, a retroflex can be an onset only if there is a preceding rhyme that ends with a vowel (Steriade).
- The *[sibilant][sibilant] filter of English has nothing to do with syllables: *[məs'ʃumiəm].
- There are also the really impressive cases that go long-distance across the word:
 - sibilant harmony
 - laterality dissimilation in liquids (cf. *lunar*, not **lunal*)
 - vowel harmony, sometime skipping across neutral vowels
- Again, the crucial matter seems to be the basis used to calculate expected frequency.

48. An Oddity

- To my knowledge, work using the observed/expected method has not used features and natural classes.
- They ought to help, because the greater generalization they achieve could get around some of the problem of “data drought” that the observed/expected method suffers from.

N

49. Observed/Expected Models

- These make direct quantitative predictions; relatively little has been done about stacking these predictions against adult intuition.
- Problem: adult intuition is evidently a blend of general principle and resemblance to particular cases (Pierrehumbert); so somehow this must be controlled for.
- It would be helpful also to gather data “backwards” in time, in hopes of seeing an evolution of intuitions reflecting further learning.

50. OT Models

- Here, theoretical development is needed: no model yet learns gradient intuitions, hampering any sort of refined assessment of the model.
- At a less grand level, the effects of particular constraints can be sought in experimental data, as in Frisch and Zawaydeh's work⁸.

51. Time for "Bake-Offs"

- There are many approaches to this problem, so it would be good to benchmark them against shared data.

⁸ Frisch, S. A. & Zawaydeh, B. A. (2001). The psychological reality of OCP-Place in Arabic. *Language* 77, 91-106.