Week 2 – Optimality Theory — Basics

February 25 and 27, 2008

1 Some problems in phonological theory that led to the invention of OT

1.1 The Conspiracy Problem

(1) Kisseberth's (1970) observation: in Yawelmani, syncope, epenthesis, allomorphy all act to create/preserve well-formed structure: all C are V-adjacent; i.e. $[CV(C)]_\sigma$

(2) Phonological rules can be seen, to some degree, as a means of forcing recalcitrant underlying forms (perhaps due to the concatenation of morphemes strings) into phonotactically well-formed surface strings.

Yawelmani: $\emptyset \longrightarrow i \;/\; C \_\_\_\; C$

(3) Rule-based theories have no concept of "phonotactically well-formed surface string," and thus cannot capture this generalization.

(4) In rule based theories, phonotactic principles are theorems of the system, deducible from morpheme structure constraints and the serially ordered rules.

(5) Colin Wilson (UCLA) envisions a "empiricist response":

a. learn the phonotactics by themselves—a topic of current work (Ellison 1992, 1994, Frisch 1996, Coleman and Pierrehumbert 1997, Frisch et al. 2004, Albright 2006, Goldsmith 2006, Hayes and Wilson 2008, Heinz 2007, Martin 2007),

b. let them be a guide to learning the alternation rules.

(i) E.g. /ʔilk+hin/ → $[[ʔi]_\sigma[li]_\sigma[khin]_\sigma]$, not faithful *$[[ʔil]_\sigma k[hin]_\sigma]$, thence knowledge that "unsyllabified C" maps onto [iC]

(ii) For work on learning phonological relations, see Gildea and Jurafsky (1996), Albright and Hayes (2002, 2003), Tesar (1995), Boersma (1997), Hayes (1999), Boersma and Hayes (2001), Prince and Tesar (2004), Hayes (2004), Tessier (2006)

(6) OT has a more direct response—essentially, the conspiracy is the phonology, as we'll see.

1.2 The International Conspiracy Problem

(7) Kiparsky (1973) "Phonological Representations".

a. Phonologies tend to aim for similar targets (avoid adjacent stresses, long consonant clusters, hiatus, rising tones, etc.).

b. But they achieve their aims in different ways–hiatus deletion (either side), gliding (either side), epenthesis of various kinds, allomorphy

(8) Rule-based theories cannot obtain a uniform characterization of the targets that the international conspiracies attempt to achieve.

a. Put in other words: rule-based theories have no explicit theory of markedness.

b. Bear in mind: not all participants think that the right approach to the "international conspiracy" problem is to load the formal theory with particular constraints.

1.3 "Do Except When" and the hierarchy of overrides

1.3.1 Axininca Campa

(9) Axininca Campa (Payne 1990): Stress falls on the 2nd, 4th, 6th syllable of a word. (We are ignoring heavy syllables and the exta complications they bring, and even-syllabled words, for now.)

[ha.ˌma.nan.ˌta.ke.ˈne.ro] 'he bought it for her'

(9) But not if it would fall on a final syllable

[ˈhi.nʲa] 'water'

(9) But yes if a monosyllable would otherwise go unstressed:

[ˈsʲ͡tsa] 'intestinal worm'

★ Write a SPE rule assigning the feature [+stress] in the above pattern.

1.3.2 Finnish Stress

(10) The Finnish "bad foot hiccup". Finnish word stress is extensively studied

- Elenbaas, Nine, and René Kager (1999), *Ternary rhythm and the lapse constraint*, Phonology 16, 273-329. `http://www.let.uu.nl/~Rene.Kager/personal/publications.htm`
- Kiparsky, Paul. 2003. Finnish Noun Inflection. In Diane Nelson and Satu Manninen (eds.) Generative Approaches to Finnic Linguistics. CSLI, 2003. `http://www.stanford.edu/~kiparsky/Papers/finnish.article.pdf`

(11) Finnish has initial main, and alternating secondaries—left to right trochees.

['jær.jes.ˌte.le.ˌmæ.tø.ˌmyː.des.ˌtæn.sæ] 'from his lack of systematization'

a. Don't stress a final syllable...

['o.pis.ˌke.li.ja] can't find gloss

b. ...unless it is the only syllable.

['puː] gloss missing

(12) Exception to the alternation pattern: if a non-initial trochee would be of the form (LH), then you make a ternary interval—the "bad foot hiccup".

['ka.las.te.ˌlem.me] 'we're fishing'
['voi.mis.te.ˌlut.te.le.ˌmas.ta] 'having caused to do gymnastics'

a. But don't make a ternary interval if you would produce final stress.

['ra.vin.ˌto.lat] 'restaurants'

(13) Finnish stress by rule

a. From left to right, form trochees. Modifications:
b. Don't make a trochee if it would be LH, but LH is in fact ok if it's final.
c. Don't make a monosyllabic trochee unless the word is monosyllabic.

(14) The ingredients of Finnish stress are utterly simple:

a. Alignment of main stress to initial syllables
b. Avoidance of final stress
c. *long lapse : three stressless syllables in a row (Elenbaas and Kager 1999)
d. Avoidance of ('LH) feet.
e. General leftward alignment of feet

(15) All of these elements can be found pervasively in stress languages, though they are probably combined in this way only in Finnish. The art of it is to prioritize them: "subject to the above ..."

2 Optimality Theory

(16) Primary original reference: Alan Prince and Paul Smolensky (1993/2004) *Optimality Theory: Constraint Ranking in Generative Grammar*, ms. published after long delay by Blackwell, Oxford. Often cited as 1993 version.

(17) Source of many other references: The Rutgers Optimality Archive: `http://ruccs.rutgers.edu/roa.html`. (This is often called the ROA.) Prince and Smolensky's original work is now posted here.

(18) Claims

a. OT can solve all three of the above problems at once.
   (i) "Rules have the effect of enforcing the phonotactics"—yes, because there

are no rules at all.
(ii) "Conspiracies involve general principles of markedness"
(iii) "Do except when" and the hierarchy of overrides

(19) The Strategy of Optimality Theory
a. State the markedness principles directly, as half of the content of the grammar
b. Other half: state principles that say: "a lexical representation can force there to be Structure X."

(20) Ranking
a. A primary (only?) source of interlanguage variation: varying ranking of conflicting principles.
b. This gives us a reasonable line on why there is such a variety of languages.

(21) The Basic Scheme
a. State markedness theory explicitly, in atomized fashion, as a large set of constraints that ban marked elements.
b. State explicit and atomized constraints that require output representations to resemble (be faithful to) input representations.
c. Let a formal procedure called GEN generate every conceivable output corresponding to the input.[1]
d. The actual output is the one that best satisfies the constraints.

(22) Result: GEN does all the derivational dirty work, and the outcome follows more or less directly from the constraints.

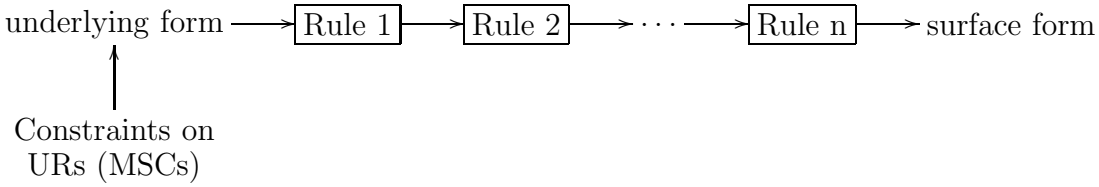

Figure 1: Traditional Generative Phonology

2.1 Defining Best Satisfaction

(23)
•Rank constraints in a hierarchy
•Recursive procedure (see Algorithm 1)

(24) Strict Constraint Domination, a.k.a. Lexicographic Order
a. Let candidate C1 violate the strictest constraint once.
b. Let candidate C2 violate constraints 2 through 1,000,000 a million times each.

[1] Gen stands for 'generator' but is always pronounced [ʤen].

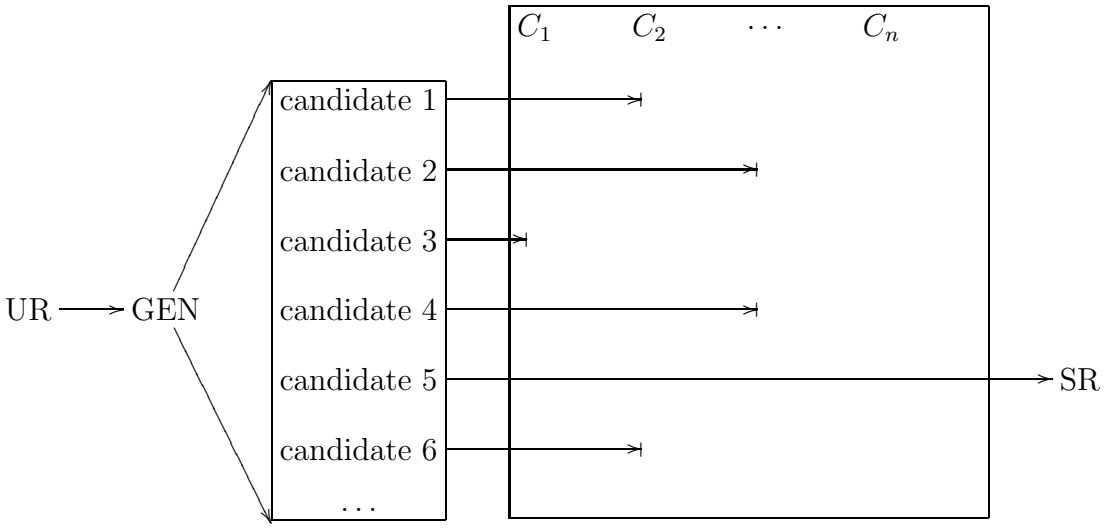

Figure 2: Optimality-theoretic Phonology

Algorithm 1 General procedure for determining the winning candidate in OT

Input: an underlying form
Ouput: the optimal candidate(s) (the surface form(s))e
With GEN, generate a set of output candidates CANDIDATES
while there are unexamined constraints **do**
1. Submit CANDIDATES to the highest-ranked constraint yet unexamined.
2. Keep only those candidates in CANDIDATES which violate this constraint *the least*. (Discard the others.)
3. Mark the highest ranked constraint as 'examined'.

end while
Designate all candidates remaining in CANDIDATES as winners (in practice there is usually one, but there could be more than one in principle in which case we need a way to interpret this, e.g. free variation).

c. The winner is C2.

(25) Analogy: the lexicographic ordering of AZZZZZZZZZZZZZZZZZ before BAAAAAAAAAAAAAAAAA

(26) This is controversial; see for instance Coleman and Pierrehumbert (1999), `http://www.phon.ox.ac.uk/~jcoleman/colepier.ps`), who attempt to refute it outright.

(27) Various methods to fudge strict domination exist, like conjoining constraints into single constraints, thus letting them "gang up."

(28) Modification for constraints that can be violated multiply: The candidate with more violations, if any, loses.

(29) Finally, note the big picture:

a. Both the traditional ordered-rule and the OT approach map URs to SRs.
b. This relation as far as is known is regular! (Johnson 1972, Kaplan and Kay 1981, 1994)
c. This does not mean we should use regular rewrite grammars (or finite state transducers), it just means we can.
d. Rather it means the focus of our investigation is this relation, which has this property (and that relations with this property have been extensively studied in other fields; however, not with natural language phonology patterns in mind).

2.2 Ranking Markedness constraints: A baby version of Finnish

(30) Miniaturizing

a. I am not fully confident that the following cases fully illustrate the system, but they hopefully will serve a pedagogical purpose. Invent some plausible constraints and rank them in a way that allows only the empirical winners to win.

H	(H) H	winner
L H	(L H) L (H)	winner
L H L L L	(L H)(L L)L L (H L)(L L) (L H)(L L)(L) (L H) L (L L)	winner
L H L H L	(L H) L (H L) (L H)(L H) L L(H L)(H L)	winner
L H L H	(L H)(L H) (L H) L H L (H L) L (L H) L (H)	winner

(31) General strategy to follow:

a. Look at the description above under §1.3.2.

b. Consider plausible constraints that rule out a lot of candidates.

2.3 There's more to Finnish...

(32) Karvonen (2005) notes an additional "hiccup" based on avoidance of ['CVC CVV] feet:

['ho.ri.son.ˌtaː.li]	'horizontal'
['sak.ra.men.ˌtaː.ri.nen]	'sacramental'
['di.ag.nos.ˌtiik.ka]	'diagnostics'
vs.	
['a.ka.ˌteː.mik.ko]	'academic'
['e.le.men.ˌtaː.ri.nen]	'elementary'
['ad.mi.nis.ˌtraːt.to.ri]	'administrator'

2.4 Folding in faithfulness

(33) Faithfulness Constraints

a. We factor the ways that two representations could differ.

b. To make the differences utterly explicit, we put an index on every segment. This is called correspondance theory (McCarthy and Prince 1995).

(34) IDENT = differ in one feature value

a. /p_1 a_2 k_3/, candidate [b_1 a_2 k_3] violates IDENT(VOICE).

(35) MAX = an underlying segment of some natural class (specified with features) is missing in the first form.

★ For UR /p_1 a_2 k_3/, what does the candidate [b_1 a_2] violate? (multiple answers)

(36) DEP = a surface segment of some natural class (specified with features) is missing in the underlying form.

★For UR /a_2 k_3/, what does the candidate [?$_1$ a_2 k_2] violate? (multiple answers)

(37) LINEARITY, violated when the linear order of any pair of segments is switched.

★ Count the violations for each of the candidates below for UR /p_1 a_2 k_3/: [k_3 a_2 p_1], [p_3 a_2 k_1], and [p_3, a_2, k_1].

(38) Not a standard Faithfulness constraint IDENT(P) "Don't change anything about [p] so it isn't [p] any more."

(39) Fundamental principle of clear analytic presentation: always declare all Faithfulness constraints violated by winners.

2.5 Defining Constraints

(40) Constraints are *functions* from (UR,SR) pairs to natural numbers.

a. IDENT(VOICE)(/p_1 a_2 k_3/, [b_1 a_2 k_3]) = 1.

(41) Note that when defining a markedness constraint, the UR doesn't matter.

a. *NOCODA(/p_1 a_2 k_3/, [p_1 a_2 k_3]) = 1
b. *NOCODA(/p_1 a_2/, [p_1 a_2 k_3]) = 1
c. *NOCODA(/p_1 a_2 k_3 a_4/, [p_1 a_2 k_3]) = 1

(42) For this reason, we can define markedness constraints as follows:

a. A markedness constraint M is a constraint such that for any SR, UR_1, and UR_2, it is the case that M(UR_1,SR) = M(UR_2,SR).

★ Why is this not true for faithfulness constraints?

(43) Fundamental principles of clear analytic presentation: whenever defining a constraint, it should be clear how to determine the number of violations for any (UR,SR) pair.

2.6 Align constraints

★ What contraints are involved in a grammar that takes pseudo-Yawelmani /patk+ma/ to patikma? Assume maximal CVC syllable structure.

(44) Prince and Smolensky (1993) suggest that Lardil forms like /kaŋ/ augment to [kaŋka] to achieve a minimal size, while aligning the stem (root+suffixes) with a syllable boundary:

	Nom (-∅)	nonfuture -in	future -uɻ	
/maɻ/	maɻt	maɻ-in	maɻ-uɻ	'hand'
/jak/	jaka	jak-in	jak-uɻ	'fish'

/maɻ/	*One-Vowel Word	*Obstruent Coda	Dep(a)	Align(Stem,R, Syl,R)	Dep(C)
☞ a. maɻ.ʈa					
b. ma.ɻa					
c. maɻ					

/jak/	*One-Vowel Word	*Obstruent Coda	Dep(a)	Align(Stem,R, Syl,R)	Dep(C)
☞ a. ja.ka					
b. jak.ka					
c. jak					

References

Albright, Adam. 2006. Gradient Phonotactic effects: lexical? grammatical? both? neither? Talk handout from the 80th Annual LSA Meeting, Albuquerque, NM.

Albright, Adam and Bruce Hayes. 2002. Modeling English past tense intuitions with minimal generalization. *SIGPHON 6: Proceedings of the Sixth Meeting of the ACL Special Interest Group in Computational Phonology* :58–69.

Albright, Adam and Bruce Hayes. 2003. Rules vs. Analogy in English Past Tenses: A Computational/Experimental Study. *Cognition* 90:119–161.

Boersma, Paul. 1997. How we learn variation, optionality, and probability. *Proceedings of the Institute of Phonetic Sciences* 21. University of Amsterdam.

Boersma, Paul and Bruce Hayes. 2001. Empirical tests of the Gradual Learning Algorithm. *Linguistic Inquiry* 32:45–86.

Coleman, John and Janet Pierrehumbert. 1997. Stochastic Phonological Grammars and Acceptability. In *Computational Phonology*. Somerset, NJ: Association for Computational Linguistics, pages 49–56. Third Meeting of the ACL Special Interest Group in Computational Phonology.

Ellison, M. T. 1992. The Machine Learning of Phonological Structure. Ph.D. thesis, University of Western Australia.

Ellison, T.M. 1994. The Iterative Learning of Phonological Constraints. *Computational Linguistics* 20(3).

Frisch, S., J. Pierrehumbert, and M. Broe. 2004. Similarity Avoidance and the OCP. *Natural Language and Linguistic Theory* 22:179–228.

Frisch, Stephan. 1996. Similarity and Frequency in Phonology. Ph.D. thesis, Northwestern University.

Gildea, Daniel and Daniel Jurafsky. 1996. Learning Bias and Phonological-rule Induction. *Computational Linguistics* 24(4).

Goldsmith, John. 2006. Information Thoery and Phonology. Slides presented at the 80th Annual LSA in Alburquerque, New Mexico.

Hayes, Bruce. 1999. Phonetically-Driven Phonology: The Role of Optimality Theory and Inductive Grounding. In *Functionalism and Formalism in Linguistics, Volume I: General Papers*, edited by Michael Darnell, Edith Moravscik, Michael Noonan, Frederick Newmeyer, and Kathleen Wheatly. John Benjamins, Amsterdam, pages 243–285.

Hayes, Bruce. 2004. Phonological acquisition in Optimality Theory: the early stages. In *Fixing Priorities: Constraints in Phonological Acquisition*, edited by Rene Kager, Joe Pater, and Wim Zonneveld. Cambridge University Press.

Hayes, Bruce and Colin Wilson. 2008. A Maximum Entropy Model of Phonotactics and Phonotactic Learning. *Linguistic Inquiry* .

Heinz, Jeffrey. 2007. The Inductive Learning of Phonotactic Patterns. Ph.D. thesis, University of California, Los Angeles.

Johnson, C. Douglas. 1972. *Formal Aspects of Phonological Description*. The Hague: Mouton.

Kaplan, Ronald and Martin Kay. 1981. Phonological Rules and Finite State Transducers. Paper presented at ACL/LSA Conference, New York.

Kaplan, Ronald and Martin Kay. 1994. Regular models of phonological rule systems. *Computational Linguistics* 20(3):331–378.

Karvonen, Daniel. 2005. Word prosody in Finnish. Ph.D. thesis, University of California, Santa Cruz. URL `http://proquest.umi.com/pqdlink?vinst=PROD&attempt=1&fmt=13&startpage=-1&ver=1&vname=PQD&did=982805101&exp=01-09-2012&scaling=FULL&vtype=PQD&rqt=309&cfc=1&TS=1168450105&clientId=1564`.

Kiparsky, Paul. 1973. Phonological Representations. In *Three Dimensions of Linguistic Theory*, edited by O. Fujimura. Tokyo: TEC Co., pages 1–135.

Martin, Andrew. 2007. The Evolving Lexicon. Ph.D. thesis, University of California, Los Angeles.

McCarthy, John and Alan Prince. 1995. Faithfulness and Reduplicative Identity. In *Papers in Optimality Theory*, edited by Jill Beckman, Laura Walsh Dickey, and Suzanne Urbanczyk, number 18 in University of Massuchusetts Occasional Papers in Linguistics. pages 249–384.

Payne, Judith. 1990. Ashéninca Stress Patterns. *Amazonian Linguistics* :185–209University of Texas Press, Austin.

Prince, Alan and Paul Smolensky. 1993. Optimality Theory: Constraint Interaction in Generative Grammar. Technical Report 2, Rutgers University Center for Cognitive Science.

Prince, Alan and Paul Smolensky. 2004. *Optimality Theory: Constraint Interaction in Generative Grammar*. Blackwell Publishing.

Prince, Alan and Bruce Tesar. 2004. Fixing priorities: constraints in phonological acquisition. In *Fixing Priorities: Constraints in Phonological Acquisition*. Cambridge: Cambridge University Cambridge: Cambride University Press.

Tesar, Bruce. 1995. Computational Optimality Theory. Ph.D. thesis, University of Colorado at Boulder.

Tessier, Anne-Michelle. 2006. Stages of OT Phonological Acquisition and Error-Selective Learning. In *Proceedings of the 25th West Coast Conference of Formal Linguistics*, edited by Donald Baumer, David Montero, and Michael Scanlon. Cascadilla Proceedings Project.