Phonotactic learning with targeted constraints

Colin Wilson, UCLA

The subset problem. In this paper, I propose a solution to the subset problem that arises for learning of phonotactic patterns of the type in (1), according to which a segment that is generally banned is nevertheless permitted as the product of assimilation. The solution makes use of the on-line error-driven version of Tesar & Smolensky's (2000) constraint demotion (CD) algorithm.

The subset problem can be illustrated by the two hypothetical languages in (2): in L1, the voiced fricative ([z]) occurs only as the result of assimilation to a following stop in L1; in L2, the voiced fricative has a less restrictive distribution (i.e., it also occurs word-initially and -finally). Positive evidence from L2 is equally consistent with both languages, because L2 is a (strict) subset of L1. How then does a learner presented only with positive examples from L2 acquire the subset language rather than the superset one?

Targeted constraints. The key idea underlying the present approach to the subset problem is that the markedness constraints that impose the relevant phonotactics are *targeted*, in the following sense:

Maximal similarity principle (MSP). Given a targeted constraint M and a candidate x that violates M, let M(x) be the set of all candidates that satisfy M better than x does. M prefers candidate y over x iff y is the member of M(x) that is most similar to x.

Building on work by Steriade (1997, 2001) and others, I take the relevant notion of similarity to be a perceptual one. This has the consequence that, relative to a candidate such as [asba], the targeted constraint against obstruent clusters that disagree in voicing--Agree(voice)--prefers only [azba], not [aspa], because the prevocalic context provides stronger perceptual cues for voicing (e.g., VOT and burst characteristics) than the postvocalic context does (see Steriade 1997 and the literature reviewed there).

The learnability result. (3) illustrates the learnability result that is achieved by replacing standard, non-targeted markedness constraints with their targeted versions. The input for each word is assumed to be the adult surface form, and demotions are made when the predicted output differs from the adult output. The initial hierarchy is one in which all markedness constraints dominate all faithfulness constraints (Levelt 1995, Smolensky 1996); therefore, initially all voiced obstruents in the input are mapped to voiceless obstruents in the output. This leads to demotion of NoVoicedStop, which prefers [pa] over [ba] and [ap] over [ab], below Ident(voice), which prefers the correct outputs. But crucially the NoVoicedFricative remains in the highest stratum--it is not demoted below Ident(voice), because it does not prefer the incorrect output [aspa] over the correct output [azba]. This follows directly from the MSP: relative to [azba], the most similar candidate that better-satisfies NoVoiFric is [asba], not [aspa].

After NoVoiStop has been demoted, the hierarchy makes the single error of mapping /azba/ to [asba]. This leads to demotion of NoVoiFric, which prefers [asba] over [azba], below the highest-ranked constraint that prefers [azba] over [asba], namely Agree(voice). At this point, the CD algorithm has returned a hierarchy that is consistent with all of the positive examples presented to it. I propose that when this point is reached the hierarchy is subject to a final operation, called *M>>F refinement* (MFR), that splits every stratum Si into two strata, one containing all the markedness constraints in Si (M-Si) and another containing all the faithfulness constraints in Si (F-Si), and ranks M-Si immediately above F-Si. The hierarchy (3d) resulting from MFR is maximally restrictive given the positive examples from L2: it allows voiced fricatives as the result of assimilation (by Agree(voice)>>NoVoicFric) but not elsewhere (by NoVoiFric>>Ident(voice)). This result generalizes to the cases in (1) with the appropriate targeted constraints, and the MFR operation introduced here can also be used to derive the initial M>>F ranking.

Theoretical consequences. The result presented obviates the additional principles that have been added to the CD learning algorithm by Hayes (1999) and Prince & Tesar (1999) in their work on the same subset problem. These principles, which primarily regulate how positional faithfulness constraints (Beckman 1998, Padgett 1995, Lombardi 1999, among others) are ranked by the algorithm, are unnecessary because targeted constraints provide a fully general alternative to positional faithfulness constraints. In the example above, the fact that Agree(voice) is targeted on the first member of the cluster in a form such as [asba] accounts for the cross-linguistic bias for regressive voicing assimilation; a positional faithfulness constraint that refers to segments in Onset or prevocalic/presonorant position is no longer needed. The present learnability result therefore converges with the results of Zoll (1998) in favoring a markedness-based approach to positional sensitivity in phonology.

Examples

(1) a. Russian (Slavic; Halle 1995). Voiced alveolar and postalveolar affricates and voiced velar fricative occur only as the result of voicing assimilation.

b. Ancient Greek (Indo-European; as suggested by Prince & Tesar 1999). Voiced fricatives occur only as the result of voicing assimilation to a following stop.

c. Modern Hebrew (Semitic; Bolozky 1978). Voiced postalveolar fricative and voiced velar fricative occur only as the result of voicing assimilation (fast speech, optional).

d. Malayalam (Dravidian; Mohanan & Mohanan 1984). Non-geminate velar nasal occurs only as the result of place assimilation to an immediately following stop.

(2) Two hypothetical distributions of stops and fricatives (based on Prince & Tesar 1999)

L1 (superset language).	pa	ap	ba	ab	sa	as	za	az	apsa	aspa	abza	azba
L2 (subset language).	pa	ap	ba	ab	sa	as			apsa	aspa		azba

(Note that 'L1' and 'L2' are arbitrary labels; they do not denote 'first language' and 'second language'.)

(3) Illustration of the learnability result: learning L2 from positive examples

a. Initial hierarchy: {NoVoicedStop, NoVoicedFricative, Agree(voice)} >> {Ident(voice)}
Mapping: /ba/ => pa /ab/ => ap /azba/ => aspa
Demotion: Demote NoVoicedStop below Ident(voice) in response to /ba/=>pa and /ab/=>ap.

b. Current hierarchy: {NoVoicedFricative, Agree(voice)} >> {Ident(voice)} >> {NoVoicedStop}
Mappings: /ba/ => ba /ab/ => ab /azba/ => asba
Demotion: Demote NoVoicedFricative below Agree(voice) in response to /azba/=>asba

c. Current hierarchy: {Agree(voice)} >> {NoVoicedFricative, Ident(voice)} >> {NoVoicedStop}
Mappings: /ba/ => ba /ab/ => ab /azba/ => azba
Demotion: None (all mappings correct).

d. M>>F refinement: {Agree(voice)} >> {NoVoicedFricative} >> {Ident(voice)} >> {NoVoicedStop}

Selected references

Beckman, J. 1998. *Positional Faithfulness*. Doctoral dissertation, University of Massachusetts, Amherst. [New York: Garland]

Bolozky, S. 1978. Some Aspects of Modern Hebrew Phonology. In R. A. Berman, ed., *Modern Hebrew Structure*. Tel Aviv: University Publishing Projects, Ltd. 11-64.

Halle, M. 1959. *The Sound Pattern of Russian*. The Hague: Mouton.

Hayes, B. 1999. Phonological Acquisition in Optimality Theory: The Early Stages. To appear in a volume on phonological acquisition and typology, edited by René Kager and Wim Zinneveld.

Levelt, Clartje. 1995. Unfaithful kids: Place of Articulation patterns in early child language. Paper presented at the Department of Cognitive Science, Johns Hopkins University, Baltimore, September.

Lombardi, L. 1999. Positional faithfulness and voicing assimilation in Optimality Theory. *Natural Language & Linguistic Theory* 17:267-302.

Mohanan, K. P., and T. Mohanan. 1984. Lexical Phonology of the Consonant System in Malayalam. *Linguistic Inquiry* 15.4:575-602.

Prince, A., and B. Tesar. 1999. Learning Phonotactic Distributions. Ms., Rutgers University, New Brunswick.

Smolensky, P. 1996. On the comprehension/production dilemma in child language. *Linguistic Inquiry* 27:720-731.

Steriade, D. 2001. The Phonology of Perceptibility Effects: the P-map and its consequences for constraint organization. Ms., UCLA.

Tesar, B., and P. Smolensky. 2000. *Learnability in Optimality Theory*. Cambridge, Mass.: MIT Press.

Zoll, C. 1998. Positional Asymmetries and Licensing. Ms., MIT, Cambridge, Mass.