Learning Long-distance Phonotactics

Heinz, 2010, a.k.a, this guy→

Presented by *Eric Sclafani*

Main idea

- Long-distance agreement (LDA) can be framed as a long-distance phonotactic pattern (LDP)
- This pattern can be learned by a **precedence learner**
- In order to generalize, the learner:
  - Takes into account the order of sounds, but not the distance between them
  - Can learn from the surface forms of words, thus not needing underlying forms
  - Does not need prerequisite phonological information such as tiers

Phonotactics: a short background

- **What is phonotactics?**
  - Rules that govern the possible valid sound sequences of a language
  - Phonotactic constraints operate over numerous properties of language such as permissible consonant/vowel clusters and syllable structure
  - Phonotactic constraints are language specific
- **Okay cool, so why study it?**
  - Evidence has shown that children learn phonotactic patterns prior to alternations.
  - Some learning models indicate that phonotactic knowledge assists with learning alternations
  - Phonotactic learning is simpler than alternations because one only has to learn a whether a string is well formed or not instead of an underlying form to a surface form.

English example:

	EXISTING	NONEXISTENT
POSSIBLE	brick	blick
IMPOSSIBLE	schwa	bnick

Source: http://seas3.elte.hu/phono/notes/141-phonotactics.html

Long-distance agreement (LDA)

- LDA are patterns in which the consonants in agreement are separated by at least one segment
- This is noteworthy because many patterns in language are local. LDA occurring between sounds that have an arbitrary distance means strictly *k*-local algorithms will not be sufficient to capture LDA patterns
- Some literature suggests that LDA is a form of feature spreading, which means a sound will “spread” one or more features to surrounding sounds
- However, feature spreading can cause blocking, but according to Hansson (2001) and Rose and Walker (2004), LDA does not have any blocking effects.
- This fact is used to show that LDA is likely not a case of feature spreading

Long-distance phonotactics (LDP)

According to the paper, we can think of phonotactic patterns as a set of well-formed strings, which acts as a function that maps strings to values. But what are these values you ask? There are two proposed approaches:

- **Categorical phonotactic model:**
  - This approach maps strings to 0 or 1 depending on if they are a member of the set of well formed strings or not.

$$f(s) = \begin{cases} 1 & \text{s is well formed} \\ 0 & \textit{otherwise} \end{cases}$$

- **Gradient phonotactic model:**
  - This approach maps strings an interval [0,1], where 0 is the “least well formed” and 1 is the “most well formed”.
  - In other words, strings are assigned probability scores.

For example, If we wanted to model English phonotactics:

- **Categorical approach**: English(slem) = 1 English(srem) = 0 English(pzarʃk) = 0
- **Gradient approach:** English (slem) = 1 English (srem) = .4 English (pzarʃk) = .1

Questions:
- Do you (dis)agree with treating the functions as categorical?
- Can you think of any potential benefits to treating them as gradient?

Long-distance phonotactics (cont.)

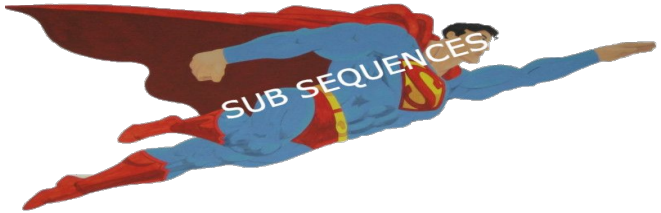
- Because phonotactic functions are also phonological ones, the question arises as to whether these functions should operate over the feature or segmental level
- However, it is unknown whether they play a role in phonotactic learnability. Thus, the paper focuses on sound segments instead of phonological features
- Additionally, whether the phonotactic patterns are treated as categorical or gradient matters little with respect to learnability. These properties don't hint at important aspects of the learning process. Therefore, in the paper, the functions are treated as categorical.

Questions:

- Do you (dis)agree with operating over the segmental level?

- Can you think of any potential benefits for looking at features instead?

Subsequences to the rescue



- Instead of solely looking at substrings, we can look at **subsequences**. Like substrings, subsequences preserve order, but not locality.
- In other words, subsequences let us look at each symbol with respect to each symbol that follows it.
- These are known as the strictly *k*-piecewise languages
- For example, the 2-piecewise subsequences of the string “murzaku” is:
  - {mu, mr, mz, ma, mk, ur, uz, ua, uk, uu, za, zk, zu, ak, au, ku}

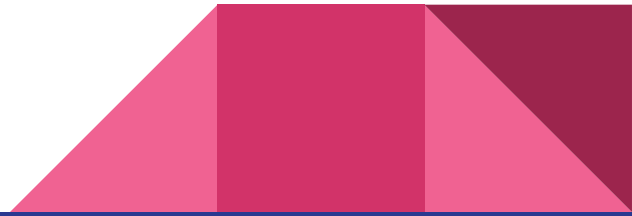
Subsequences continued

- Surprise! This technique can be used to also capture LDP!
- For example, one could write the following constraints for Navajo, where the anteriority of sibilants in a word is influenced by the rightmost sibilant:
  - ***s . . . ʃ**
  - ***s . . . ʒ**
  - ***[α anterior] . . . [-α anterior]**
- With these rules, *sotos* and *toʃotoʃ* are valid surface forms, but *sotoʃ* and *ʃotos* are not.

a. /sì-ʔá/ → sì-ʔá ‘a round object lies’
b. /sì-tí/ → sì-tí ‘he is lying’
c. /sì-ɣìʃ/ → ʃì-ɣìʃ ‘it is bent, curved’
d. /sì-teːʒ/ → ʃi-teːʒ ‘they (dual) are lying’

Precedence Grammars and Languages

- A **precedence grammar** is defined as the class of strictly 2-piecewise languages and any language generated by this grammar is called a **precedence language**.
- In this case, the term **precedence** refers to a relation between symbols in some string. If symbol **x** and **y** stand in a precedence relation in string **s**, then **xy** is a subsequence of **s.**



Precedence grammars and languages (cont.)

(17) In well-formed words, sibilants must agree in the feature [anterior].

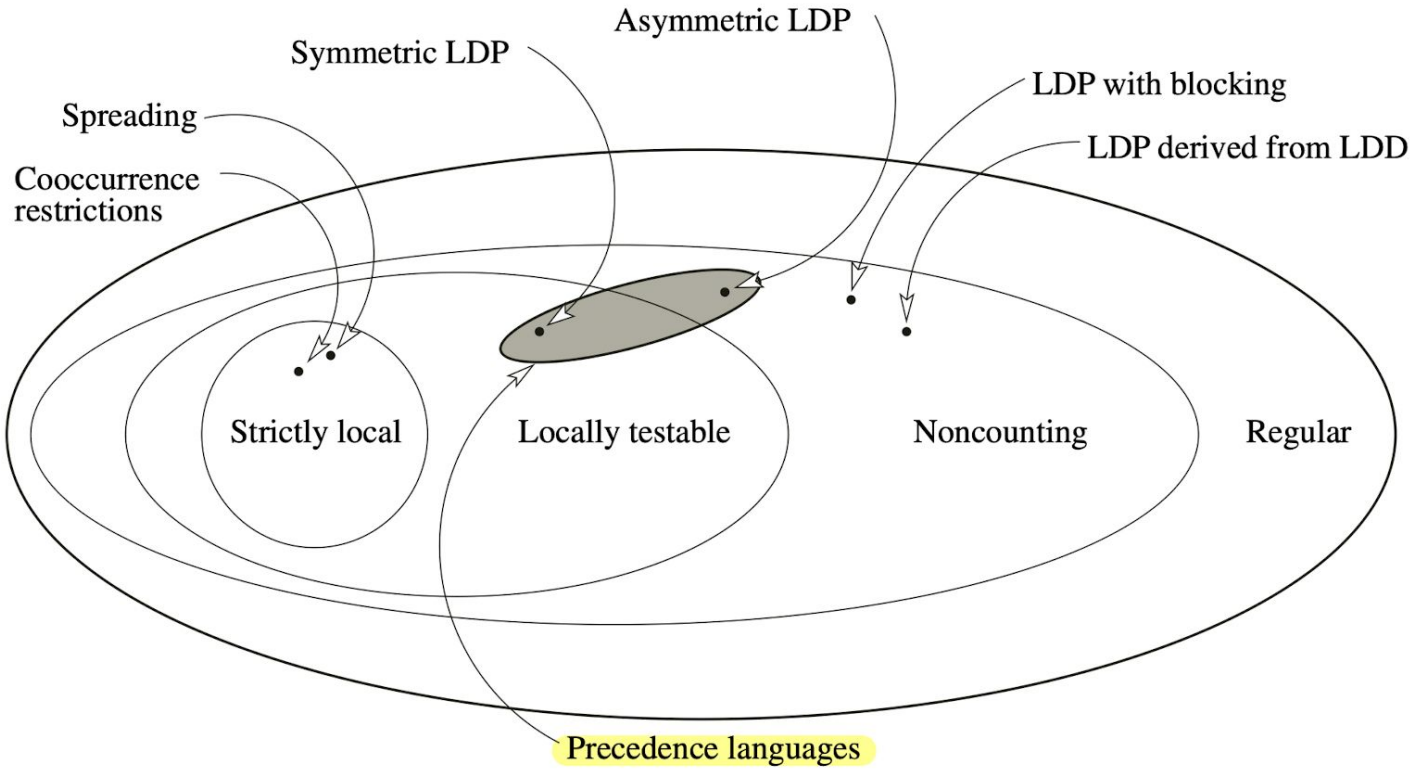
(18) 1. [−anterior] sibilants never precede [+anterior] sibilants.
2. [+anterior] sibilants never precede [−anterior] sibilants.

(19) [s] can be preceded by [s].
[s] can be preceded by [t].
. . .
[t] can be preceded by [s].
. . .
[ʃ] can be preceded by [ʃ].
[ʃ] can be preceded by [t].
. . .

Table 6
A precedence grammar for a fragment of Navajo

$$G = \begin{Bmatrix} ss & & st & so \\ & ʃʃ & ʃt & ʃo \\ ts & tʃ & tt & to \\ os & oʃ & ot & oo \end{Bmatrix}$$

Precedence grammars and languages (cont.(cont.))



Towards a phonotactic learning model

- The **precedence learner** is evaluated in the Gold (1967) *language Identification in the limit* framework because it focuses more so on generalization from positive data alone.

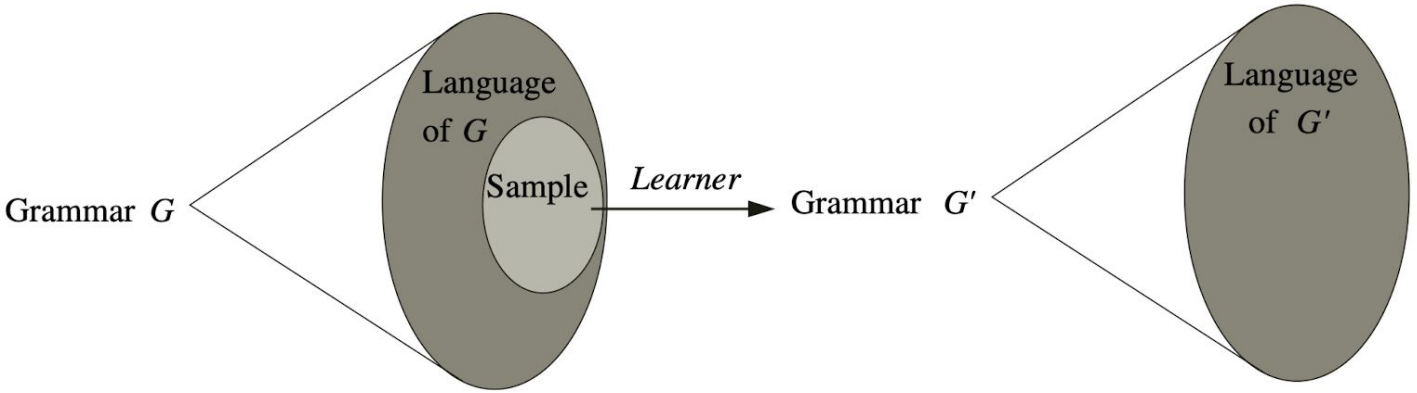


Figure 4
The learning framework

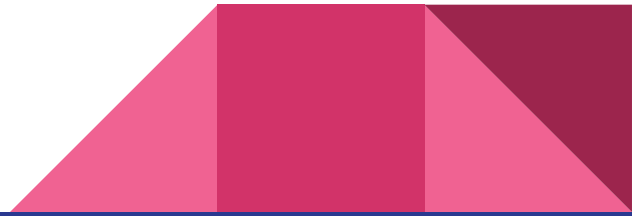
Towards a phonotactic learning model (cont.)

Table 7
Precedence learning: Navajo sibilant harmony

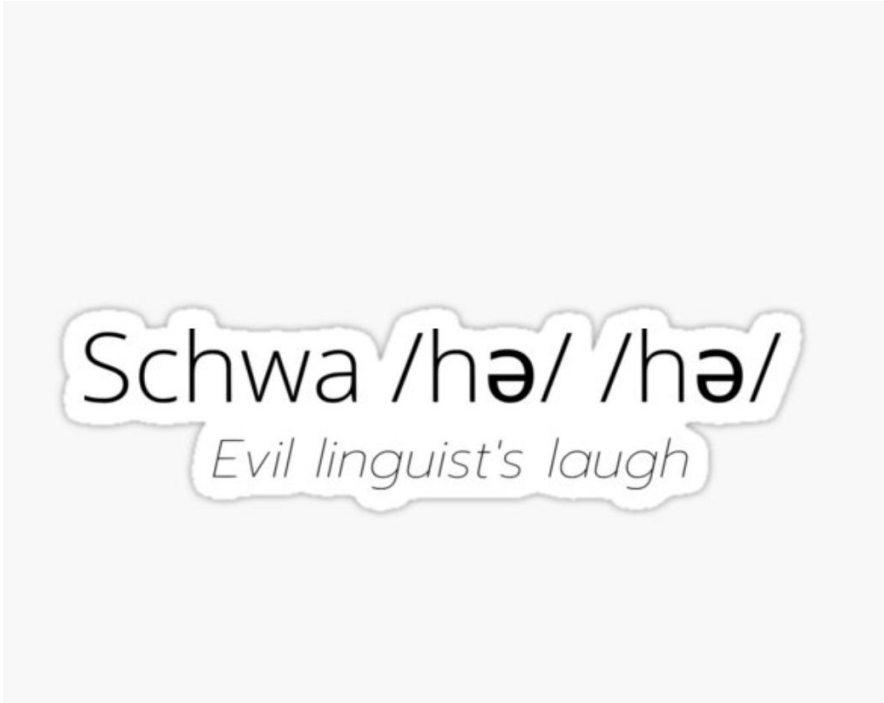
Time	Word	Precedence relations	Grammar
0			Ø
1	tosos	to, ts, os, oo, so, ss	{**ss**, **so**, **ts**, **to**, **os**, **oo**}
2	ʃotoʃ	ʃo, ʃt, ʃʃ, ot,oo, oʃ, to, tʃ	{ss, **ʃʃ**, **ʃt**, **ʃo**, ts, **tʃ**, to, os, **oʃ**, **ot**, oo}
3	stot	st, so, to, tt, ot	{ss, **st**, so, ʃʃ, ʃt, ʃo, ts, tʃ, **tt**, to, os, oʃ, ot, oo}

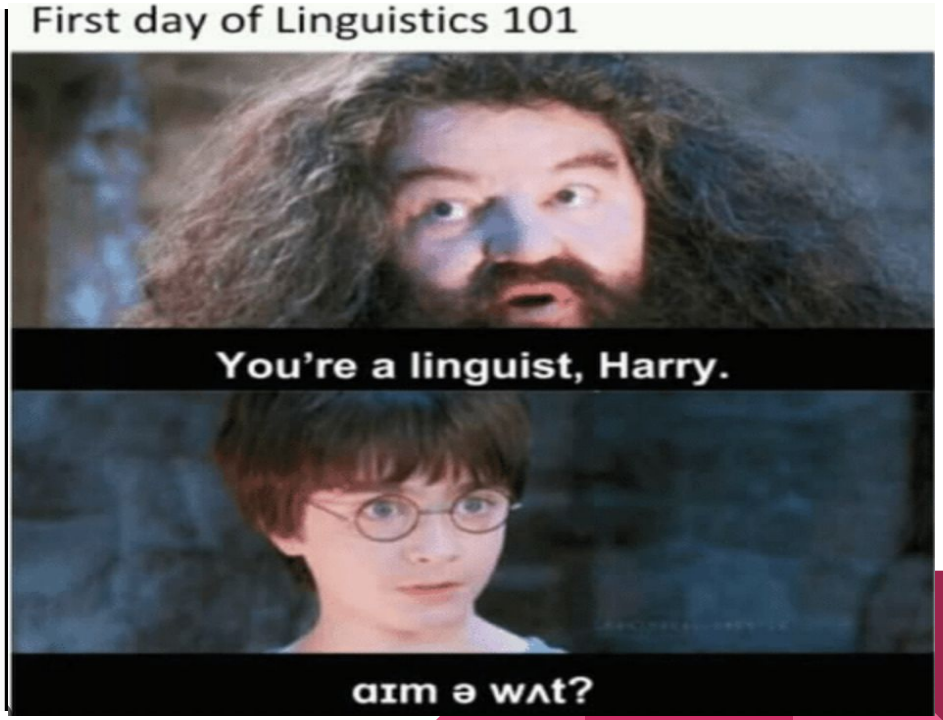
Modular Language Learning

- Just employing a subsequence learner is not enough, however.
- What happens when a nonsense word contains valid subsequences?



FIN~


Schwa /hə/ /hə/
Evil linguist's laugh


First day of Linguistics 101
You're a linguist, Harry.
aɪm ə wʌt?