Deep Phonotactics
(a multiply ambiguous title)

Stefan A. Frisch

University of South Florida

Research supported by NIH NIDCD,
Fulbright Scholar Program, and USF

The Big Picture

- Aren’t statistical phonotactics just frequency tracking and outside the grammar?
- Phonotactic patterns are tied to higher levels (language, prosodic structure) and lower levels (segments, phonetic cues) and there’s no justification to separate out the categorical ones from the statistical ones

The Big Picture

- Static phonotactics in the lexicon provide an opportunity to observe weak/gradient/gradual influences
  - Soft influence of functionality can alter the lexicon over time
  - Influence should be seen in all levels
  - Primary focus: Case study examining consonant clusters

Probabilistic Phonotactics

- The statistical distribution of phonological constituents in the lexicon
  - Based on type frequency (dictionary frequency) rather than token frequency (usage) (Bybee)
- Influences language processing in a variety of ways
  - Phoneme frequency or neighborhood density effects in perception and production (Luce, Vitevitch)
  - Metalinguistic judgments of well-formedness of novel nonwords (Bailey & Hahn 2000)

Deep$_1$ Phonotactics?

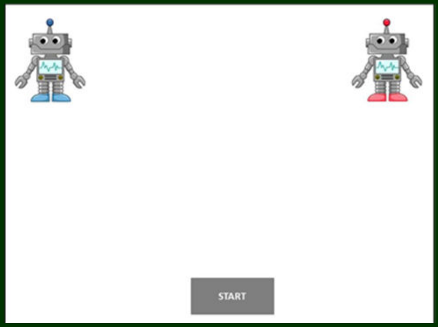
- There are statistical patterns in language, known by language speakers
- Is this just frequency tracking?
  - – Frequencies are useful for language learning (constructing categories)
    - Parsing words from speech (e.g. Saffran)
    - Assigning word internal phonological (e.g. Treiman et al 2000) or morphological constituents (e.g. Hay)
    - Language specific information for bilinguals

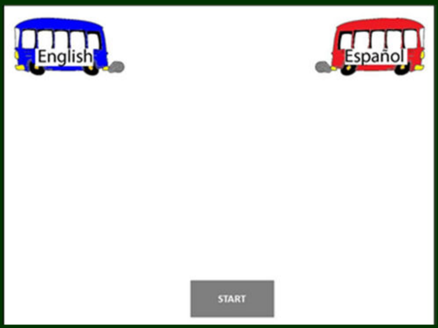
Bilingualism & Phonotactics

- Young bilingual children know statistical phonotactic patterns, in both of their languages (Messer, Leseman, Boom, & Mayo, 2010; Sebastián-Gallés & Bosch, 2002)
- Information used to parse two languages in computational models (e.g. Li & Farkas, 2002; Shook & Marian, 2013)

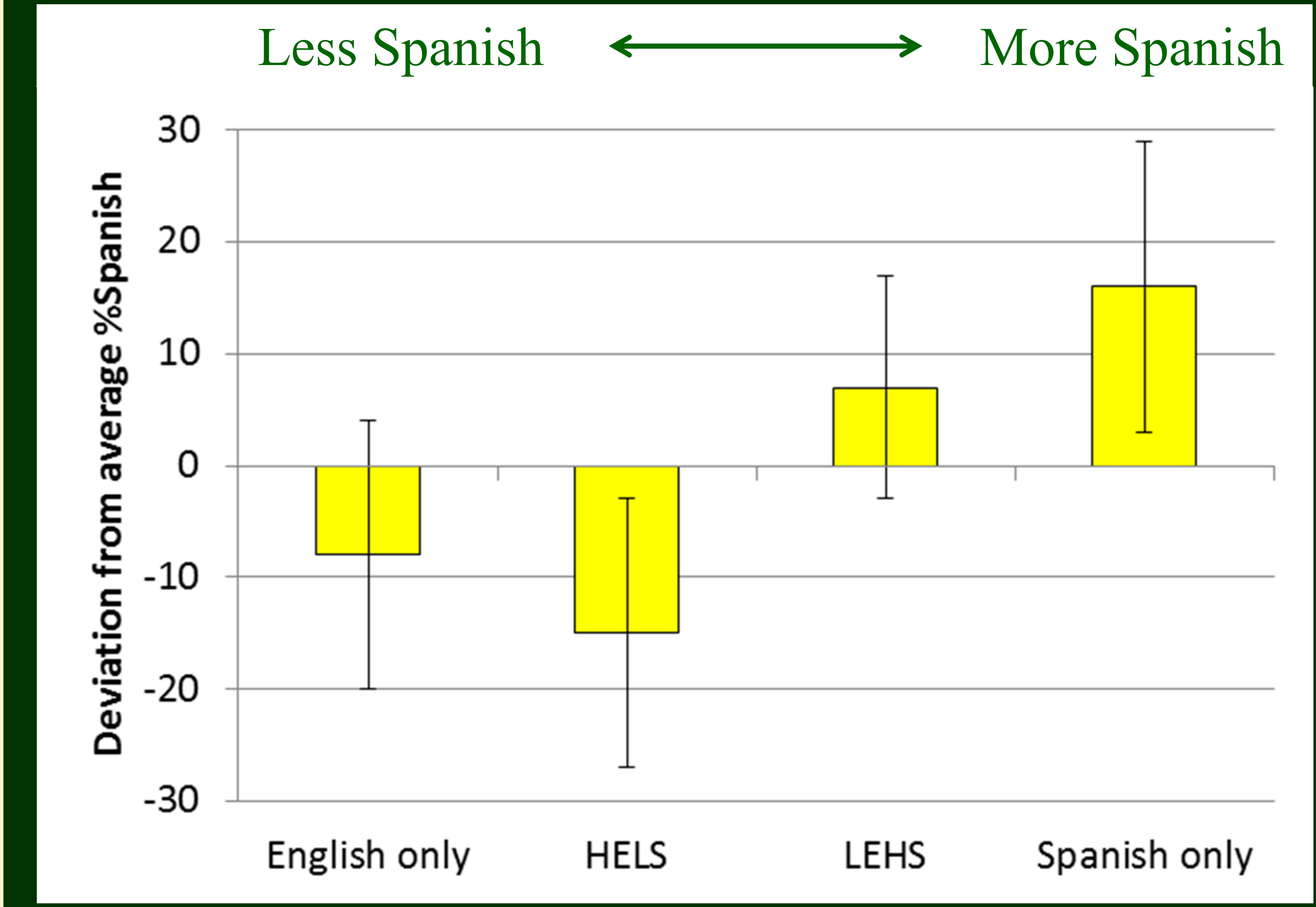
Betancourt data

- Betancourt (2013) found metalinguistic judgments by 5 year old Spanish-English bilingual children can be biased by phonotactics
  - Resynthesized/morphed stim
  - Phonetics neutralized
  - Associated with robots
  - Spanish or English speaking?





Betancourt data



Phonological Constraints

- Early probabilistic analysis of the lexicon: Greenberg’s (1950) study of consonant co-occurrence in Arabic
- Recent attempts to model statistical details (Anttila 2008, Coetzee & Pater 2008, Frisch et al 2004, Hayes & Wilson 2008, Alderete et al 2013)
- In other words, recent analyses implement a phonological constraint that is probabilistic

Functional Explanation

- Arabic constraint is to avoid repeating similar consonants at the same place of articulation within a word (OCP-Place)
- Repetition may be especially difficult to process given templatic morphology (Berg 1999)
- Less strict (even more probabilistic) versions found in any language where someone has looked

Modeled by Freq Tracking?

- Alderete, Tupper, & Frisch (2013) trained a connectionist network with a hidden layer on the Arabic root lexicon
  - – Consonants represented with features
  - – Restricted size of the hidden layer forced generalization across segments
  - – Target of learning was to classify consonant sequences as attested (or not)

Modeled by Freq Tracking?

- Results of c-net modeling
  - Network learned roots (lower acceptability for novel roots after training)
  - Network encoded OCP (lower acceptability for OCP violating roots, novel or attested)
  - Similarity influence on OCP found for novel roots used in a metalinguistic experiment (Frisch & Zawaydeh 2001)

Modeled by Freq Tracking?

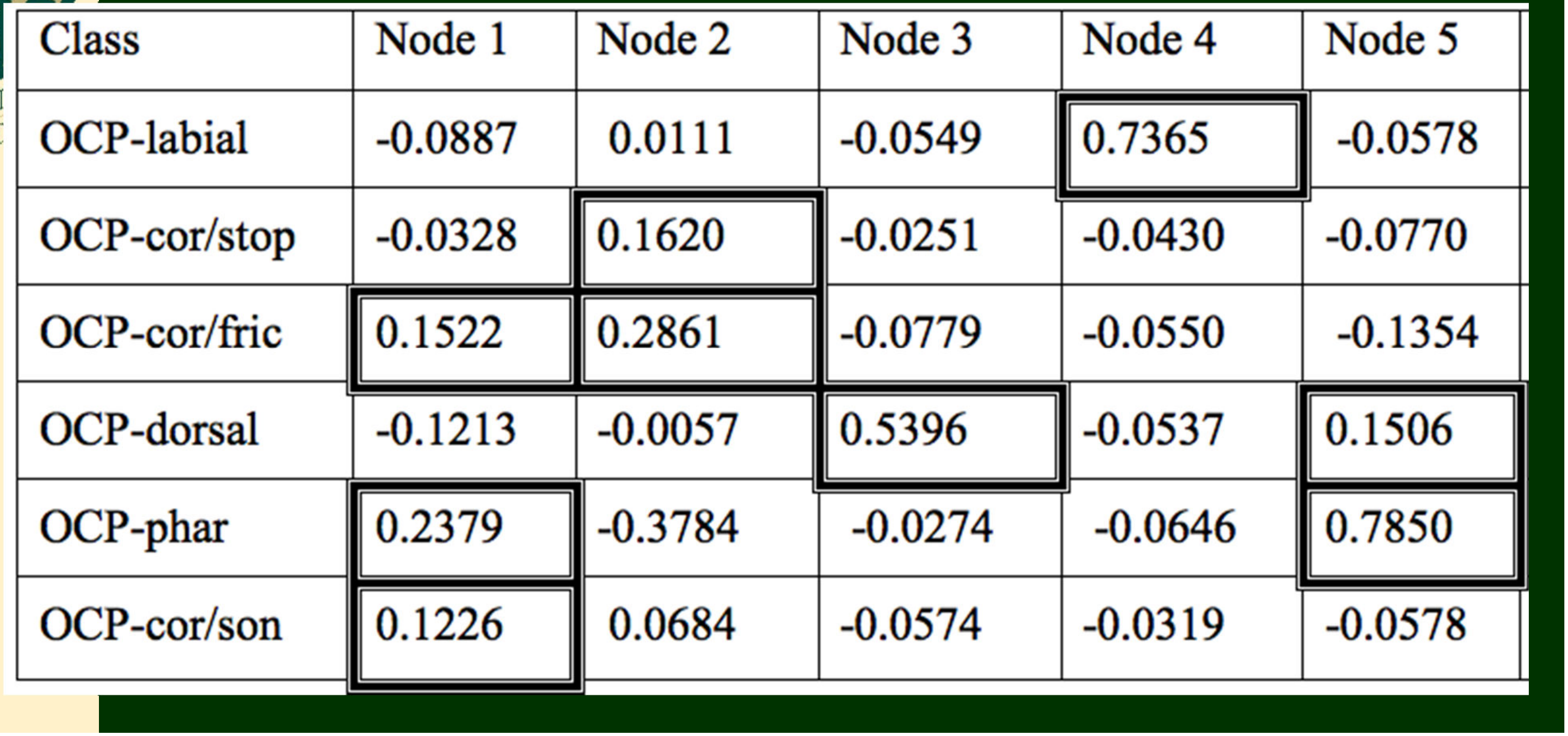
- Frequency tracking using distributed representation and backpropogation learning in a connectionist network (Hebbian-type learning)
- Different from recent phonological grammar learning models in that the phonological part is all in the representation (but cf. Hayes constraint deduction)
- The deduced constraints in Alderete et al (2013) are distributed (but potentially discoverable)

Analysis of the Model

- Correlation of hidden node activation with violation of OCP
- Some place classes represented in one node, others in two
- Often, strength of correlations and overlaps are interesting (i.e. reflect a pattern in the data)

Analysis of Model

- Labial stands alone strongly, coronal & dorsal more overlapping, coronal weak
- Consistent with lexical data



Class	Node 1	Node 2	Node 3	Node 4	Node 5
OCP-labial	-0.0887	0.0111	-0.0549	0.7365	-0.0578
OCP-cor/stop	-0.0328	0.1620	-0.0251	-0.0430	-0.0770
OCP-cor/fric	0.1522	0.2861	-0.0779	-0.0550	-0.1354
OCP-dorsal	-0.1213	-0.0057	0.5396	-0.0537	0.1506
OCP-phar	0.2379	-0.3784	-0.0274	-0.0646	0.7850
OCP-cor/son	0.1226	0.0684	-0.0574	-0.0319	-0.0578

Interim summary

- Statistical phonotactic patterns are connected to linguistic categories at various levels
- Statistical patterns can be learned by frequency tracking

Diachrony

- Difficult to process patterns may be at an evolutionary disadvantage (Blevins 2004)
- Forces acting on the sound patterns of language should be found throughout all levels of the system
- As an analogy, water erosion can produce macro level structures like canyons as well as micro level structures like small rivulets in the canyon walls. The large canyon is an accumulation of rivulets acting over time.

Deep$_2$ Phonotactics

- Difficult to process patterns should be discoverable in a variety of ways typologically
  - Common occurrence of traditional phonological constraint (Statistical trend across languages)
  - Statistical trends within languages
    - In the possible constituents they have (number of accidental gaps)
    - In the frequency with which those constituents occur (in the lexicon)

Evidence for Functionality

- Strongest evidence for an evolutionary perspective if evidence for a functional force are found at all levels
- Can be seen as evolutionary forces acting throughout the system as language is used (Blevins, also Bybee)

Present Study

- Exploring this kind of evidence in sonority constraints for consonant clusters
  - Greenberg previously provided a statistical cross-linguistic typological result for sonority sequencing – sonority rise toward the peak
  - Ohala argues for sonority modulation – sonority changes are good

Functionality of sonority

- Sonority sequencing may be the result of ease of production (jaw opening/closing cycle)
- Sonority modulation may be the result of ease of perception (alternating high/low dominant frequency spectra makes segments easier to identify)
- Both of these constraints would work against having lots of clusters

Stochastic sonority constraints

- Some overlap with beats-and-binding work on sonority in consonant clusters (Dzubalska-Kolaczyk)
- Data looking at model ranking of cluster types vs. type/token frequency
- Also similar in spirit to Baroni (yesterday) examining a variety of clusters across languages

Stochastic sonority constraints

- Quantitative lexical typological study
  - Analysis of consonant cluster type frequency for 47 languages
  - Initial, medial, and/or final clusters
  - Compare the occurring CC clusters within a language to possible clusters (given types of C1 and C2 that are used in clusters)
  - Focus on CC for ease of analysis and higher frequency of this size cluster

Stochastic sonority constraints

- Example: Abun
  - – Word initially, Abun only allows C-glide clusters
  - – Abun has 7 stops, 4 prenasalized stops, 3 fricatives, 3 nasals, and 2 glides
  - – Abun has 11 different C-glide clusters, mostly of the type stop-glide
- Metric: A/P
  - – Actual # of clusters of each type
  - – Possible # of consonant combinations

Abun C-glide clusters

C1	Actual	Possible	A/P
stop	6	7x2 = 14	0.43
presnas stop	3	4x2 = 8	0.38
fric	1	3x2 = 6	0.17
nas	1	3x2 = 6	0.17

- Abun pattern follows the predictions of sonority sequencing if we look at the number of cluster types

More Complex Example

English coda cluster types		C2			
		stop	fric	nas	liq
C1	stop	0.06	0.10		
	fric	0.06	0.02		
	nas	0.22	0.25		
	liq	0.92	0.69	0.67	0.25

Stochastic sonority constraints

- Metric: Compare actual vs. possible clusters in adjacent cells (e.g. stop-stop vs. stop-fricative)
  – Sonority sequencing predicts more frequent stop-fricative for codas, less frequent fricative-stop for onsets, and is neutral for medial clusters
  – Sonority modulation predicts stop-stop less frequent than stop-fric regardless of position
  – Analysis will start with modulation

Example

Frequency should increase for modulation

English coda clusters		C2			
		stop	fric	nas	liq
C1	stop	0.06	0.10		
	fric	0.06	0.02		
	nas	0.22	0.25		
	liq	0.92	0.69	0.67	0.25

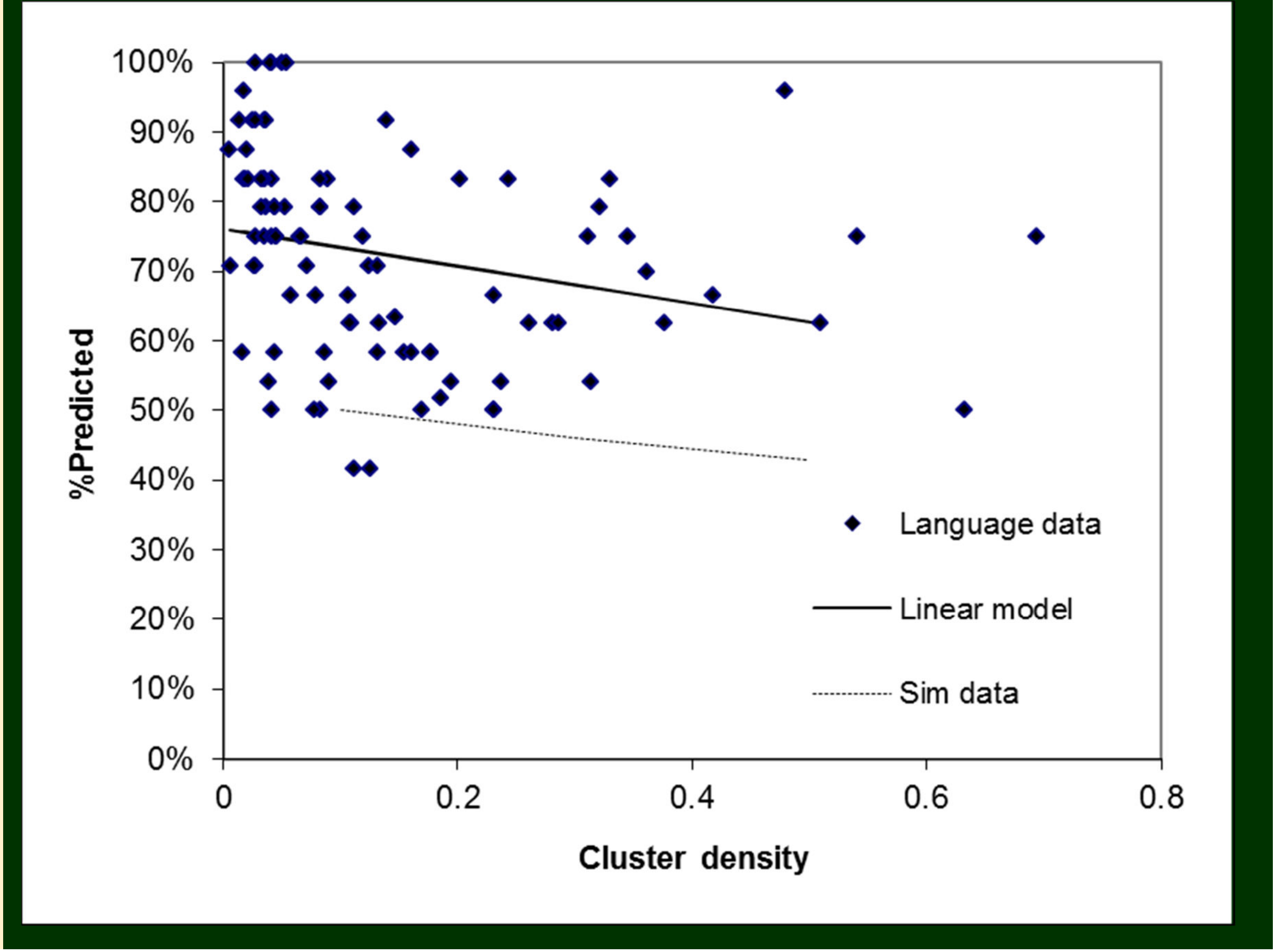
Basic Finding

- Modulation obeyed in 72% of comparisons over 93 cluster data sets (over 47 languages)
- Tendency toward sonority modulation
- Caveats:
  - One-step pairwise comparison measure
  - Data semi-overlapping (up-down/left-rt)
  - Multiple cluster data sets within a language probably not independent
  - What should we expect by chance? 50%?

Interactions

- Sonority modulation can be in conflict with sonority sequencing (sonority fall in an onset or sonority rise in a coda)
- Density of clusters in the language (treated as an independent factor) provides pressure against sonority – if many clusters, can't all be optimal
- Monte Carlo simulations for comparison at 0.1, 0.3, 0.5 density

Sonority Modulation & Density



UNIVERSITY OF SOUTH FLORIDA

Interactions

- Sonority sequencing and sonority modulation are in conflict in roughly half of the data table
- Can look at influence of sonority sequencing by comparing influence of this conflict on success of predictions of sonority modulation

Sonority Modulation & Sequencing

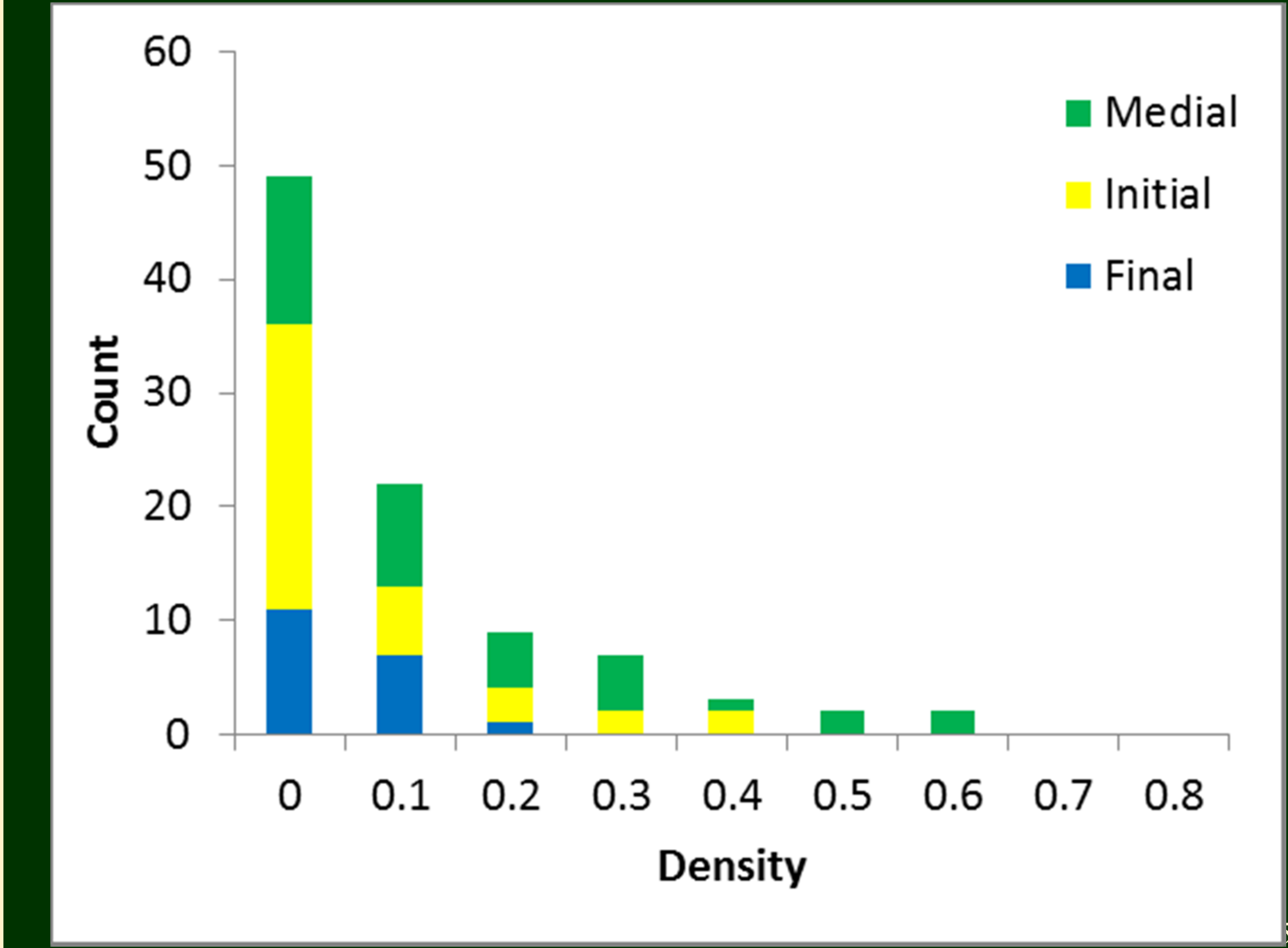
		Constraints agree		Constraints conflict	
Position	N	Predicted	SD	Predicted	SD
Initial	38	0.76	0.15	0.69	0.16
Final	18	0.79	0.11	0.58	0.24
Medial*	37	0.71	0.16	0.71	0.18

*no constraint conflict in medial clusters, but data were still split as though they are onsets (onset bias)

Summary: Cluster Type Freq

- More “gaps” are found in permissible sonority combinations when the sonority difference is small compared to when the sonority difference is large
- Even more so when the sonority difference violates sequencing
- Harder to obey sonority if the language uses many clusters (in fact, constraint against density of clusters “stronger”?)

Cluster Density in Sample



Cluster lexical frequency

- Functional sonority constraints would predict the avoidance of sonority applies as individual words are used
- Evolutionary approach predicts words with better clusters should also be more commonly found within a language (cluster lexical frequency)

Case Study: Spanish Medial

Spanish medial count		C2				
		stop	fric	nas	liq	glide
C1	stop	299	385	70	1214	520
	fric	1522	137	162	515	1852
	nas	2575	931	43	16	328
	liq	578	610	308	17	528
	glide	93	92	65	37	0

Case Study: Spanish Medial

Spanish medial O/E		C2				
		stop	fric	nas	liq	glide
C1	stop	0.24 → ↓	0.77 → ↑	0.47 → ↑	1.97 → ↑	0.69 ↑
	fric	0.94 ← ↓	0.21 → ↓	0.82 → ↑	0.63 → ↑	1.87 ↑
	nas	1.74 ← ↓	1.54 ← ↓	0.24 → ↓	0.02 → ↑	0.36 ↑
	liq	0.80 ← ↓	1.76 ← ↓	2.94 ← ↓	0.04 → ↓	1.01 ↑
	glide	0.86 ←	2.07 ←	4.87 ←	0.68 ←	0.00

Results

- 68% of comparisons obey modulation
- Similar to cluster type frequency
- Soft constraint against poor sonority clusters acting on the lexicon

Semi-Big Picture

- Like the OCP, sonority constraints should be found, at least statistically, everywhere we look
- Over time functional forces shape the phonological system (what is attested or not) leading to categorical constraints in some languages and typological tendencies across all languages

Conclusion

- Deep Phonotactics
  - In an evolutionary phonology perspective, functional constraints on language processing should affect patterns at all levels (phonotactics a great place to look)
  - But, statistical patterns an important foundation of linguistic knowledge, emergent from learning the lexicon and language usage (mapping from experiences to categories)

USF
UNIVERSITY OF
SOUTH FLORIDA

USF
UNIVERSITY OF
SOUTH FLORIDA

Thank you.