SONORITY SEQUENCING IN POLISH: DEFYING THE STIMULUS?

GAJA JAROSZ

FORMAL APPROACHES TO SLAVIC LINGUISTICS 25
MAY 13-15, 2016, ITHACA, NY

HOW DO WE LEARN PHONOTACTIC GENERALIZATIONS?

- Phonological knowledge encodes general preferences
  - ‘mip’ > ‘bwip’ > ‘dlap’ > ‘bzap’
- Where do these preferences come from?
- Lexicalist Hypothesis: Phonotactics derived from lexical statistics
  - Dominant hypothesis in phonology & language acquisition
  - Incontrovertible evidence for lexical frequency
    - Learning & processing
- Universals: what shapes languages?
  - To what extent does learning shape/derive universals?
- Unresolved debate: how constrained is LAD?
  - Constraints on representations?
  - Constraints on frequency sensitivity?
  - Behavioral results often consistent with input statistics and universals
- Today: Statistics vs. Universals
  - Syllable structure phonotactics
  - Predictive computational models can help

NATURE VS. NURTURE

- Increasingly "Universalist" Hypotheses
  - Raw statistics, no generalization/similarity
    - Analogy (Bailey & Hahn 2001)
    - Phoneme co-occurrence (Vitevich & Luce 2004)
  - Frequency Sensitivity with Class-Based Generalization (CBG)
    - Includes various abstract representations: features, syllables, tiers, etc.
    - UCLA Phonotactic Learner (Hayes & Wilson 2008)
    - Featural Bigram Model (Albright 2009)
  - Universal bias and frequency sensitivity
    - Preferences among abstract representations: VoicedCoda < VoicelessCoda
    - Innate grammatical principle
    - Universal phonetic pressure
- This talk
  - Review evidence concerning roles of Universals and Statistics
  - Discuss evidence that at least CBG is needed
  - Present novel evidence supporting Universal Bias
    - Focus: modeling syllable structure phonotactics in Polish

STATISTICS: DEVELOPMENT

- Infant learning is exquisitely sensitive to statistics
  - 9 month olds sensitive to gradient phonotactics (Jusczyk et al. 1993)
  - 8 & 9 month olds can use various statistical regularities to segment fluent speech (Saffran et al. 1996; Mattys & Jusczyk 2001; and many more)
  - Phonotactic probability affects production/perception of contrasts in later development (Zamuner et al. 2004, Zamuner 2009, Zamuner 2013)
- Different acquisition orders in different languages
  - /v/ earlier in Swedish, Estonian, & Bulgarian than English (Ingram 1988)
  - /l/ early in French but late in English (Vihman 1993; Edwards & Beckman 2008)
  - Earlier acquisition coincides with higher relative frequency <u>in that language</u>

STATISTICS: ADULTS

- Gradient acceptability
  - Coleman & Pierrehumbert (1997); Bailey & Hahn (2001); Vitevich et al. (1997); Frisch et al. (2000); Hayes & Londe (2006); Zuraw 2000; Ernestus & Baayen 2003; Becker et al. 2011
- Various tasks
  - Non-word repetition (Vitevich & Luce 1998)
  - Speech errors (Goldrick & Larson 2008; Goldrick 2011; Levitt & Healy 1985; Motley & Baars 1975)
  - Picture naming (Levelt & Wheeldon 1994)
  - Short-term memory recall (Gathercole & Martin 1996)
  - Phoneme Identification (Pitt & McQueen 1998)
- Lexicalist computational models capture behavioral data
  - Hayes & Wilson (2008), Albright (2009), Daland et al. (2011), Coetzee & Pater (2008); Hayes & Londe (2006), Bailey & Hahn (2001), Vitevich et al. (1997)

STATISTICS: MODELS

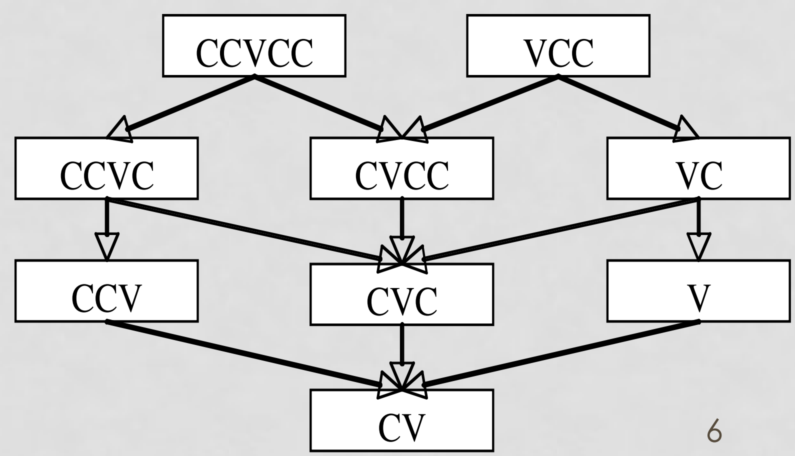
- UCLA Phonotactic Learner (Hayes & Wilson 2008)
  - English onsets phonotactic judgments (from Scholes 1966)

Model	*r*
Our model	0.946
Clements and Keyser 1983 constraints with maxent weights	0.936
Coleman and Pierrehumbert 1997	0.893
Our model without features	0.885
N-gram model	0.877
Analogical model	0.833

- This is a lexicalist model:
  - It constructs constraints for under-represented patterns
    - *[+son,+dor] - no dorsal nasals
    - *[+son][ ] - no sonorant-initial clusters
  - Weights constraints to match lexical patterns

UNIVERSALS: DEVELOPMENT

- Strongest arguments for universal constraints
  - Children show sensitivity to principles / constraints not evident in the input
- Universal restrictions on development
  - Development respects <u>universal implicational hierarchies</u>
    - CV before CVC, CV before V, CVC before CVCC, etc…
    - <u>Cross-linguistically</u>: Dutch, German, English, French, Polish (Lleo & Prinz 1996; Levelt et al. 2000; Kirk & Demuth 2005; Demuth et al. 2006; Jarosz 2010; Jarosz et al. submitted)



UNIVERSALS: ADULT GRAMMARS

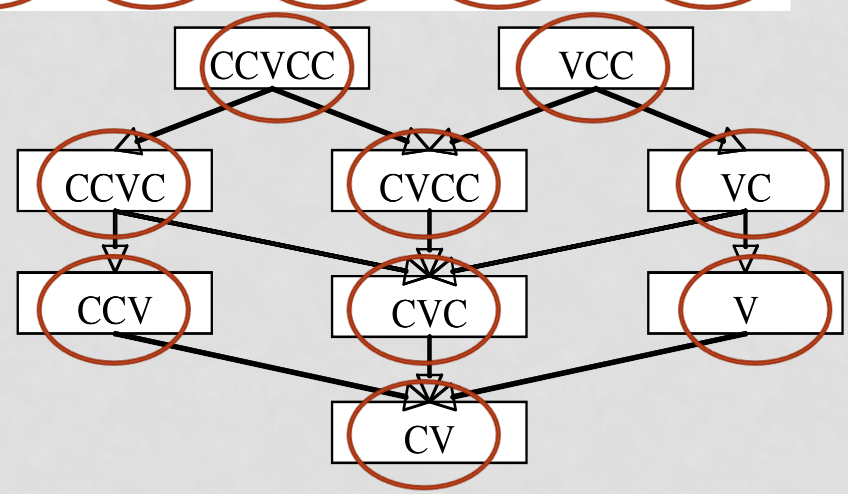
- **Poverty** of the Stimulus Arguments
  - Production/perception/acceptability of illicit forms
    - Illegal clusters in English (Berent et al. 2007, Berent 2008, Berent & Lennertz 2009, Berent et al. 2009, Davidson et al. 2004, Davidson 2006)
    - Illegal clusters in Mandarin, Korean (Berent et al. 2008; Ren et al. 2010)
  - Second language acquisition, loan adaptation
    - Devoicing in Mandarin L2 acquisition of English codas (Broselow et al. 1998)
    - Sonority effects in Hindi, Japanese & Korean L2 acquisition of English onset clusters (Broselow & Finer 1991; Eckman & Iverson 1993)
- **Surfeit** of the Stimulus Arguments
  - Turkish (Becker et al. 2011) and English (Becker et al. 2012) speakers fail to generalize statistical regularities predictive of laryngeal alternations
  - Unnatural constraints don't predict phonotactic acceptability judgments in English (Hayes & White 2013)
  - Speakers underlearn unnatural regularities predictive of vowel harmony (Hayes et al. 2009)

STATISTICS MIRROR UNIVERSALS

- Universals are often mirrored by language particular statistics!
  - Polish relative frequencies (Jarosz 2010)

CV	CVC	CVCC	V	VC	VCC	CCV	CCVC	CCVCC
50·3%	20·9%	3·3%	8·5%	3·5%	0·6%	9·6%	4·0%	0·6%

  - When marked structures are allowed, they tend to be under-represented



CHICKEN OR EGG?

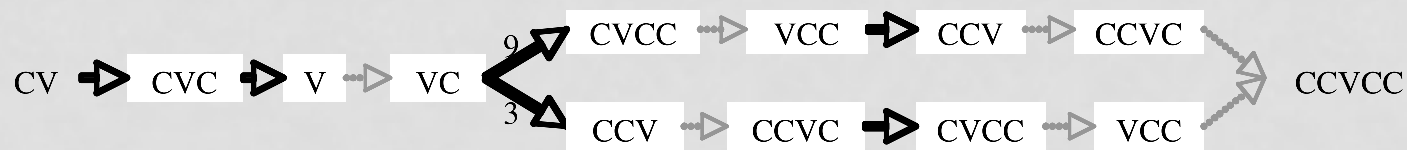
- Statistics mirror universals
  - Many results consistent with both
  - Problem 1
    - Hard to tease apart
  - Problem 2
    - Circular: how did input come to resemble the universals?
- Next: Progress with Computational Modeling
  - Generate <u>explicit, testable predictions</u>
  - Identify <u>crucial test cases</u>
  - Modifications lead to <u>theory development</u>

MODELING: PROGRESS

RAW STATISTICS?

STATISTICS REFERENCE STRUCTURE

- Raw statistics worked in Polish example
- Other cases: statistics have to operate over abstract features and classes
- Dutch Acquisition Order (Levelt & van de Vijver 2000)



- Dutch input frequencies

CV	CVC	CVCC	V	VC	VCC	CCV	CCVC	CCVCC
44·8%	32·1%	3·3%	3·9%	12·0%	0·4%	1·4%	2·0%	0·3%

- Frequency-sensitivity must be structure-sensitive
  - Holistic: f(VC) > f(V), f(CCVC) > f(CCV)
  - Combinatorial – independent components:
  - f(…V#) = 50.1 f(#CV…) = 80.2
  - f(…VC#) = 46.1 f(#V…) = 16.3
  - f(...VCC#) = 4.0 f(#CCV…)= 3.9

FEATURE-BASED GENERALIZATION

- UCLA Phonotactic Learner (Hayes & Wilson 2008)
  - English onsets phonotactic judgments (from Scholes 1966)

Model	*r*
Our model	0.946
Clements and Keyser 1983 constraints with maxent weights	0.936
Coleman and Pierrehumbert 1997	0.893
Our model without features	0.885
N-gram model	0.877
Analogical model	0.833

- This model crucially references natural classes
  - Models without performed worst
  - Other work with similar findings
    - Albright (2009), Daland et al. (2011), Coetzee & Pater (2008)

OVERVIEW

- Increasingly "Innatist" Hypotheses
  - Raw statistics, no generalization/similarity
  - Frequency Sensitivity with Class-Based Generalization
  - Something more (innate principle, phonetic pressures…)
- Pushing CBG to its limits
  - Identify cases where Lexicalist models & Universals conflict
  - Use modeling to disentangle the predictions
- Focus: SSP
  - SSP is part of syllable structure phonotactics
  - Long tradition of typological work on universals
  - Recent modeling, experimental progress

SONORITY SEQUENCING PRINCIPLE

- **Sonority Sequencing Principle** (SSP; Clements 1988, Selkirk 1984)

  [lb]ack < [nb]ack < [bd]ack < [bn]ack < [bɹ]ack < [bj]ack

  -2 -1 0 1 2 3

- **Sonority Scale** (amplitude)
  - Glides: 4: [j, w]
  - Liquids: 3: [l, ɹ, r]
  - Nasals: 2: [m, n, ɲ, ŋ…]
  - Obstruents: 1: [p, t, k, b, d, g, f, s, v, z, x, h…]
- **Typological Findings**
  - Syllable inventories (Berent et al. 2007; Greenberg 1978)
    - [bn]ack ⇒ [bɹ]ack
  - Syllabification
    - English: rug.by vs. u.gly

SONORITY PROJECTION

- **Sonority Sequencing Principle** (SSP; Clements 1988, Selkirk 1984)

[lb]ack < [nb]ack < [bd]ack < [bn]ack < [bɹ]ack < [bj]ack

[lb]ack	[nb]ack	[bd]ack	[bn]ack	[bɹ]ack	[bj]ack
-2	-1	0	1	2	3

- Consistent findings of **Sonority Projection** in English
  - Preferences between unobserved clusters
    - **#nb** (-1) vs. **#db** (0)
  - Documented using various tasks
    - Production, perception, acceptability; aural, written (Berent et al. 2007, Berent & Lennertz 2009, Berent et al. 2009, Davidson et al. 2004, Davidson 2006, Daland et al. 2011)
- Question:
  - Where do these preferences come from?
    - Statistics are 0 for all
  - Could they be learned?

ENGLISH: POVERTY OF THE STIMULUS?

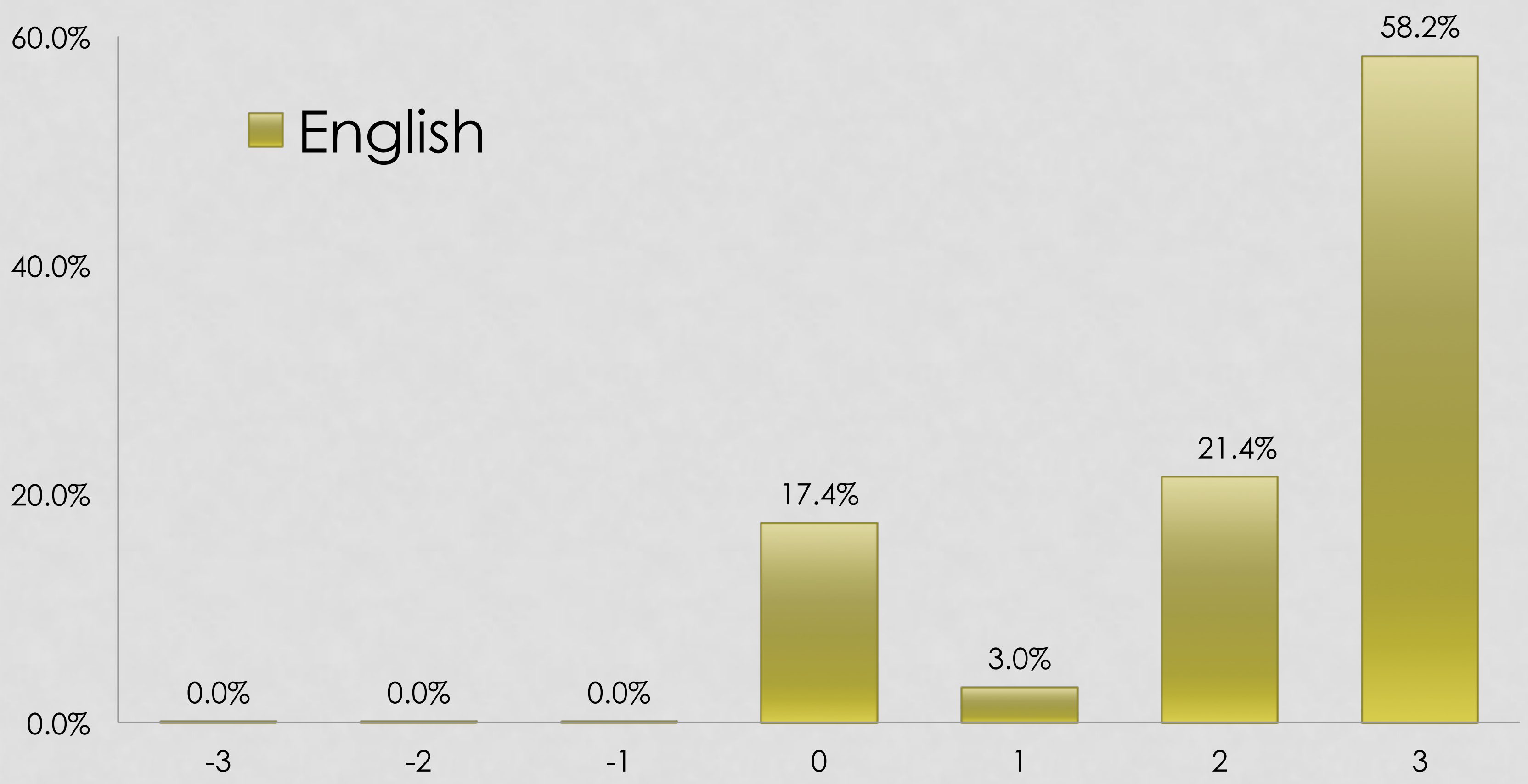
- Berent et al. (2007): Poverty of the Stimulus
  - English speakers exhibit sonority projection effects
    - *[lb]ack (-2) < *[bd]ack (-1) < *[bn]ack (1)
  - Basic/raw lexical statistics of English don't capture effect
- Daland et al. (2011): No Poverty of the Stimulus
  - Several lexicalist models derive SSP for English
  - These models can capture experimental results
  - They do not need to encode SSP preference
  - As long as they have:
    - **Syllable structure** tells the model [gb] in rug.by doesn't count
    - **Features** tell the model what sounds are similar to one another
    - **Frequency Sensitivity** allows models to favor more frequent patterns
      - *#[+son][-son] vs. *#[-son][+son]
      - Captures preference for #[**bn**]ack over #[**nb**]ack
      - More words in English similar to [bn] than to [nb]
- Raw Statistics not sufficient

ENGLISH SYLLABLE ONSETS

OO (0)		ON (1)		OL (2)		OG (3)	
st	521	sn	109	fl	290	pr	1046
sp	313	sm	82	kl	285	tr	515
sk	278			pl	238	kr	387
				bl	213	gr	331
				sl	213	br	319
				gl	131	fr	254
						dr	211
						kw	201
						sw	153
						hw	111
						θr	73
						tw	55
						ʃr	40
						dw	17
						gw	11
						θw	4
(17.4%)		**(3.0%)**		**(21.4%)**		**(58.2)%**	

(data from Hayes & Wilson 2008)

ENGLISH



MORE IMPOVERISHED STIMULUS?

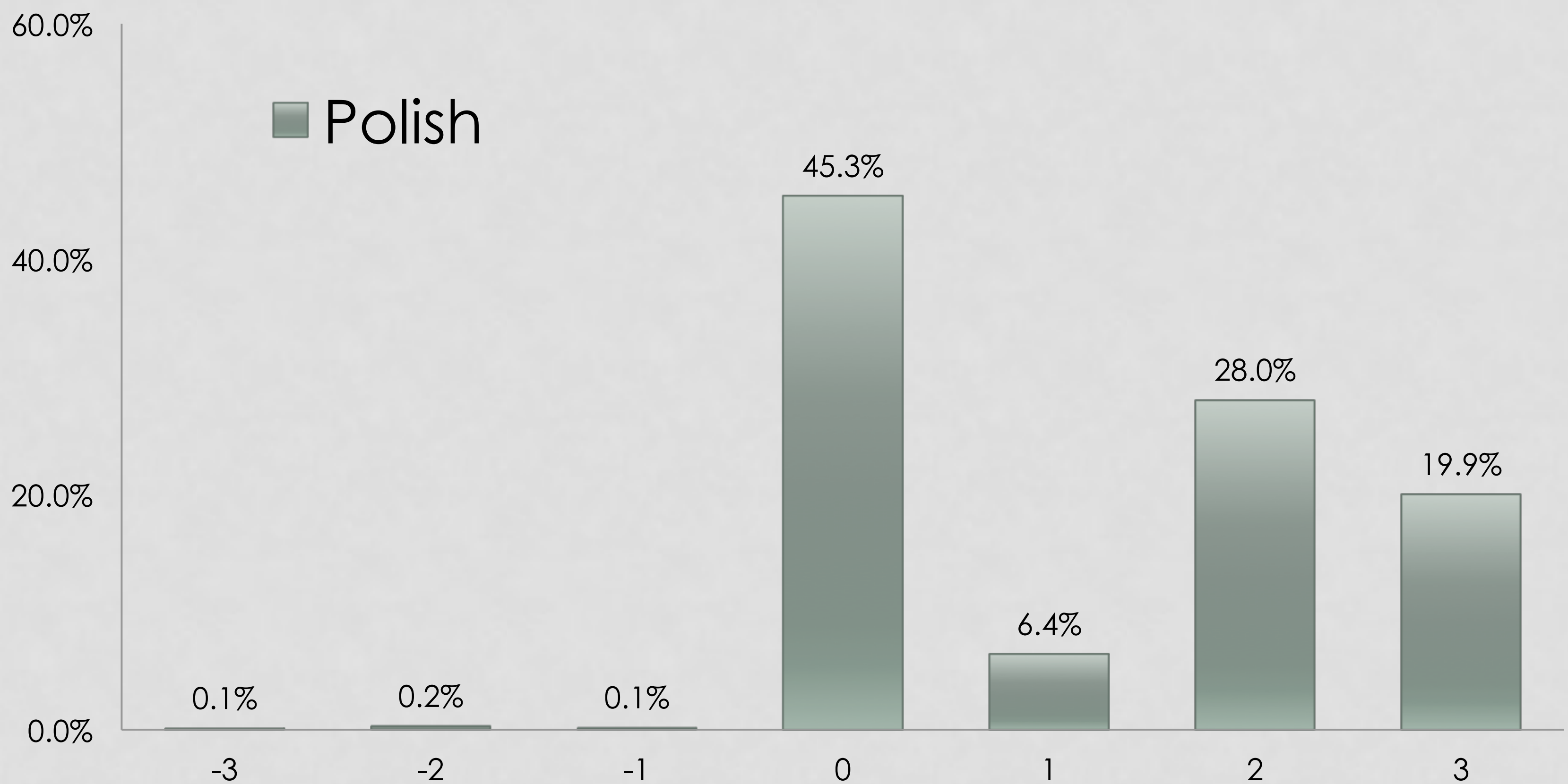
- More impoverished stimulus?
  - Korean and Mandarin arguably lack onset clusters
    - Though there are obstruent-glide-vowel sequences…
  - Korean (Berent et al. 2008) and Mandarin (Ren et al. 2010) speakers also exhibit sonority projection effects!
- Hayes (2011) on Ba and Bwa language
  - UCLA Phonotactic Learner supports SSP generalization!
    - Given UG with bigram constraints on sonority features
  - Lots of #[-son][+son] and similar from Ba and Bwa
  - No #[+son][-son] or similar
  - Similar predictions for *any* language with only rises!
- At an abstract level, these languages mirror SSP!
  - These are not the strongest test cases for lexicalist models
  - Need: language input that doesn't mirror SSP

POLISH?

THE OPPOSITE: POLISH?

- What does the input look like in Polish?
  - Whole Scale!
  - [wb]ack < [lb]ack < [mb]ack < [bd]ack < [bn]ack < [bɹ]ack < [bj]ack
  - -3 -2 -1 0 1 2 3
  - [wza] [lvɨ] [mʂa] [ptak] [dnɔ] [kluʧ] [zwɨ]
- Polish Child Directed Speech Sample
  - From Polish CDS Frequency Dictionary (Haman 2011)
    - ~800k word tokens (~115k #CC)
    - ~44k word types (~11k #CC)

THE OPPOSITE: POLISH?



OBSTRUENT-OBSTRUENT CLUSTERS

- All combinations are well attested
- Traditional analyses note lack of sequencing restrictions among obstruents (Rubach & Booij 1990a; 1990b)

	FP	PP	FF	PF
Frequency	1910	140	510	2378
Examples:	st	kt	sx	pʂ
	sp	tk	xf	kɕ
	sk	db	fx	ps
	zb	pt	vz	kʂ
	zd	gd	zv	bʑ
	zg		ʂf	tʂ
	vd			gʐ

SSP IS ACTIVE IN POLISH

- SSP is active in **<u>Comparative Allomorphy</u>** (Rubach 1986; Bethin 1987; Rubach & Booij 1990a, 1990b)
  - [ʂ] is added to most adjectival stems:
    - /swab + ɨ/ 'weak' → [swap + ʂ + ɨ] 'weaker'
    - /zdrɔv + ɨ/ 'healthy' → [zdrɔf + ʂ + ɨ] 'healthier'
    - /mwɔd + ɨ/ 'young' → [mwɔt + ʂ + ɨ] 'younger'
    - /star + ɨ/ 'young' → [star + ʂ + ɨ] 'younger'
    - /tfard + ɨ/ 'hard' → [tfart + ʂ + ɨ] 'harder'
  - [ɛjʂ] is added to avoid <u>Obs-Son-Obs</u> sequences:
    - /wadn + ɨ/ 'nice' → [wadɲ + ɛjʂ + ɨ] 'harder'
    - /ʂt͡ʂupw + ɨ/ 'thin' → [ʂt͡ʂupl + ɛjʂ + ɨ] 'thinner'
  - [ɛjʂ] is variably added to avoid <u>Obs-Obs-Obs</u> sequences (contra Rubach):
    - /watf + ɨ/ 'easy' → [watfj + ɛjʂ + ɨ] 'easier'
    - /t͡ʂɨst + ɨ/ 'clean' → [t͡ʂɨstɕ + ɛjʂ + ɨ]~[t͡ʂɨst + ʂ + ɨ] 'cleaner'
    - /gɛw̃st + ɨ/ 'thick' → [gɛw̃stɕ + ɛjʂ + ɨ]~[gɛw̃st + ʂ + ɨ] 'thicker'
- SSP is relevant to **<u>Voicing Processes</u>** (Rubach & Booij 1987, 1990, 1990b)
  - Final obstruent devoicing: /kadr/ → [katr]
  - Regressive voicing assimilation: /jabwkɔ/ → [japwkɔ]

QUESTIONS AND PREDICTIONS

- Is the SSP principle derivable from the input?
- Sonority Falls are really rare
  - Universal SSP predicts late development
  - Statistics predict late development
  - Some evidence this is right
    - Łukaszewicz (2006) shows 3;8-5;1 child has OO clusters, but not Son-O clusters
- Focus: Plateaus & Rises
  - Predictions of Universal SSP
    - OO, NN (0) < ON, ML (1) < OL, MG (2) < OG (3)
  - Predictions of Lexicalist Models
    - Development sensitive to input statistics

DEVELOPMENT

THE ACQUISITION DATA

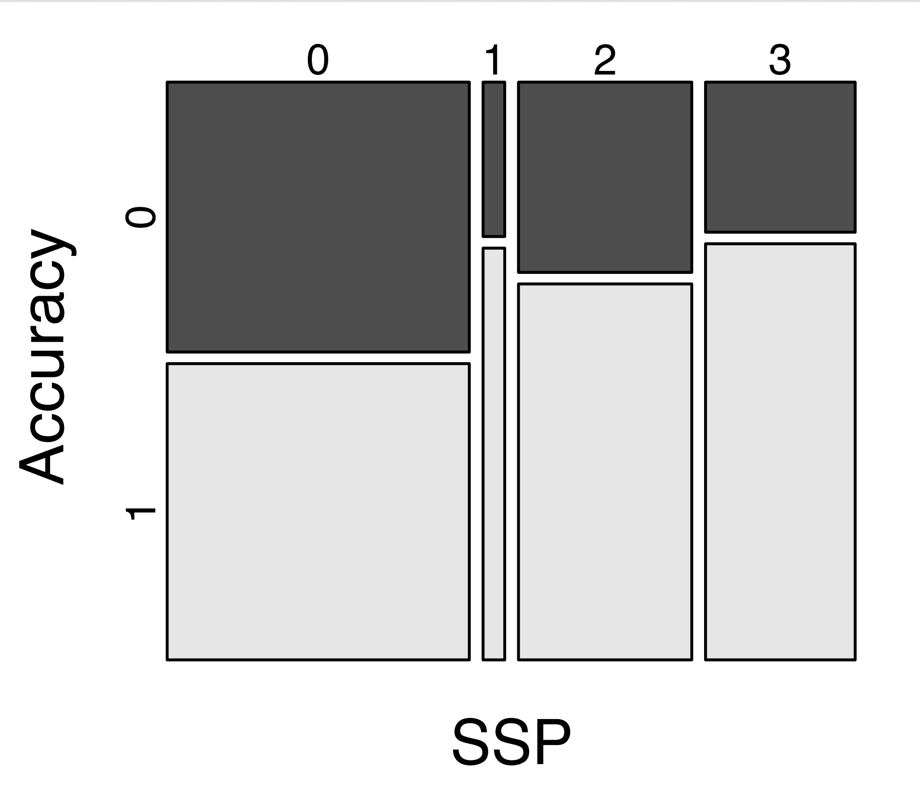
- Spontaneous Production Data
  - Weist-Jarosz Corpus

    (Weist et al., 1984; Jarosz 2010; Jarosz et al. 2015)
  - 4 Children in two age groups
    - 1;7-1;11 2;1-2;6
  - Broad phonetic transcription
    - Aligned Actual and Target → accuracy
    - ~9000 word tokens

Bartosz (6 transcripts)	Kubuś (5 transcripts)	Marta (3 transcripts)	Wawrzon (4 transcripts)
Ages 1;7-1;11	Ages 2;1-2;6	Ages 1;7-1;8	Ages 2;2-2;6
2192 words	2772 words	1218 words	2913 words

ACCURACY BY SSP

- Raw data visualization
- Focus on Plateaus & Rises
  - Children didn't attempt falls
- Accuracy rises with SSP



ACQUISITION: ANALYSIS

- Logistic Regression: predicting accuracy on initial clusters
- Dependent Variable: **Accuracy**
  - Coded at sonority level: unrelated substitutions (Place) don't affect results
- Factor of interest: **SSP**
- Statistical **Controls**
  - Age
  - Length of target word in syllables
  - Log Word Frequency
  - Primary Stress
  - Participant
  - Function word
  - Morphologically complex

REGRESSION RESULTS

	β	S.E.	Wald Z	Pr(>|Z|)
Intercept	-11.16	1.92	-5.8	**<0.0001**
Subject				
Kub	9.24	2.66	3.47	**0.001**
Mar	9.86	6.02	1.64	0.101
Waw	10.48	1.94	5.4	**<0.0001**
Age	0.47	0.09	5.38	**<0.0001**
Function Word	-0.29	0.17	-1.64	0.100
log(Word Frequency)	-0.11	0.03	-3.51	**0.001**
Stress	0.33	0.20	1.64	0.101
Prefix	-0.65	0.14	-4.47	**<0.0001**
Word Length	0.05	0.12	0.45	0.650
SSP	**0.28**	**0.04**	**7.16**	**<0.0001**
Kub * Age	-0.38	0.11	-3.46	**0.001**
Mar * Age	-0.43	0.30	-1.4	0.163
Waw * Age	-0.42	0.09	-4.79	**<0.0001**

ACQUISITION SUMMARY

- Main Question: Is SSP predictive?
  - **Yes**
  - Higher sonority rise predicts higher accuracy
- Same conclusion with P<F<N<L<G<V scale
  - Nested comparisons
    - P<F<N<L<G<V subsumes O<N<L<G<V
      $(\chi^2(1) = 0.25, p > 0.61)$
  - Support for P<F?
    - Preference for P<F is not evident in target language
      - (Rubach & Booij 1990a; 1990b)
    - But children seem to be sensitive

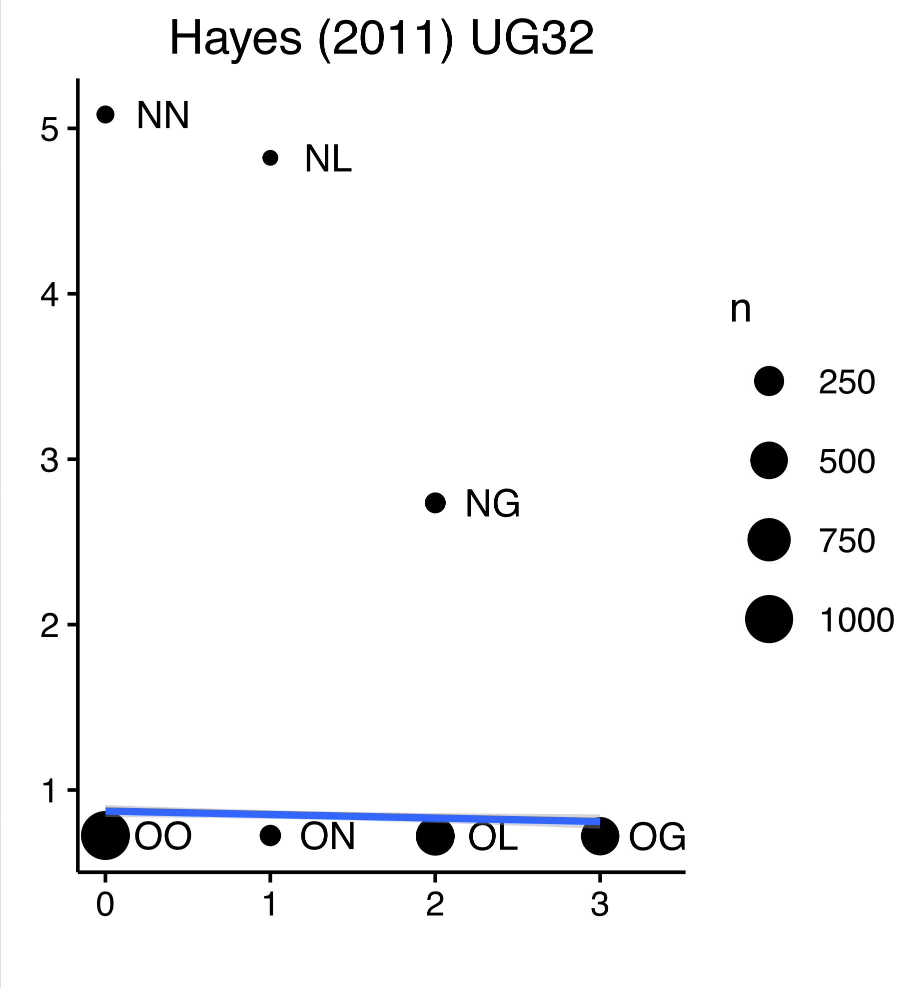
EVALUATING LEXICALIST MODELS

UCLA PHONOTACTIC LEARNER

- UCLA Phonotactic Learner (Hayes & Wilson 2008)
  - One of most successful models in previous work
  - It was able to detect SSP in English (inducing constraints)
  - It was able to detect SSP in toy Mandarin/Korean (fixed constraint set)
- Testing/Training
  - Training: all onsets in target language
  - Testing: all possible onset clusters
- Examined both ways
  - Given constraints (like for toy Mandarin/Korean)
  - Inducing constraints (like for English)

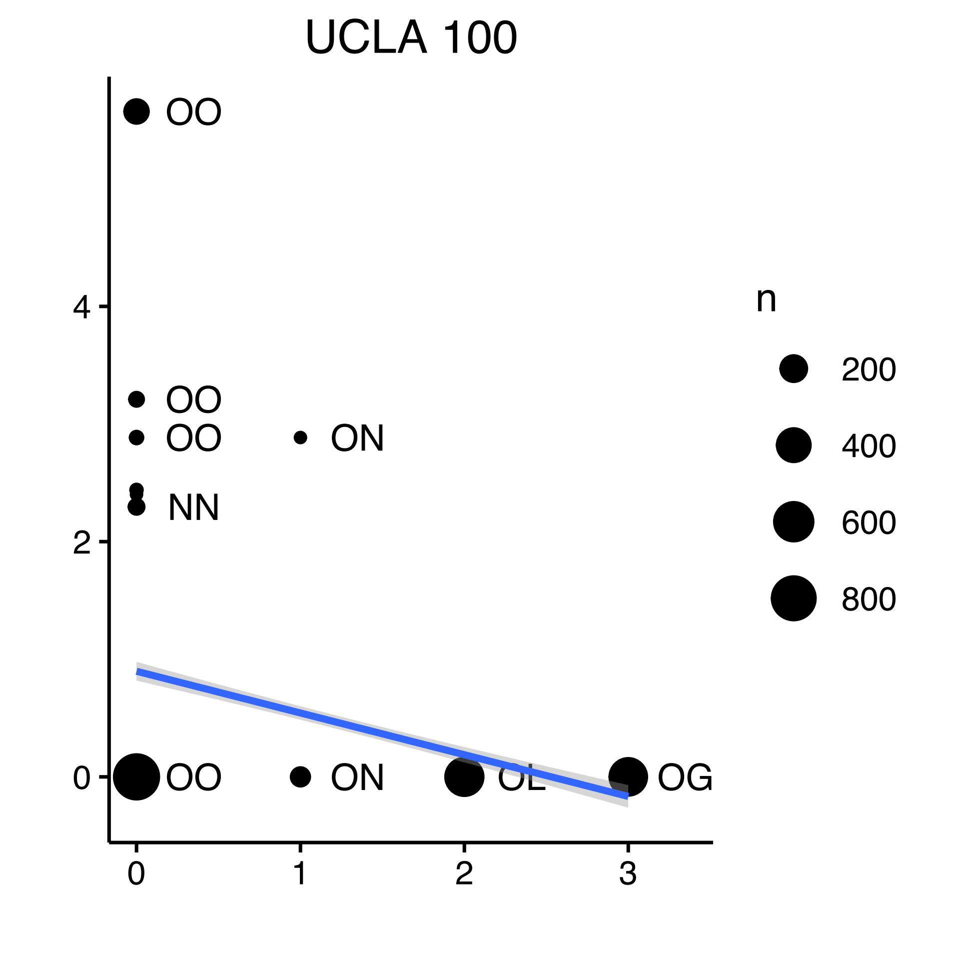
UCLA PHONOTACTIC LEARNER: PRE-SPECIFIED CONSTRAINT SET

- Model assigns **penalties**
- Successful sonority projection
  - Higher penalties for lower rises
- No sonority projection!
  - >90% clusters get same score=0



UCLA PHONOTACTIC LEARNER: INDUCING CONSTRAINTS

- There is a trend
  - Due to singled-out clusters
  - [ʣv], [vʥ], [mɲ]…
- Again, no penalties for most data
- No sonority-related penalities



MODEL COMPARISONS: SSP PREDICTIVE

	Induce 100		Induce 200		Hayes2011 UG32		Hayes2011 UG64	
	+F	+F+SSP	+F	+F+SSP	+F	+F+SSP	+F	+F+SSP
D_{xy}	0.453	0.473	0.447	0.468	0.427	0.46	0.435	0.459
LR	408.48	437.88	394.43	427.41	360.19	412.81	371.97	412.26
$\chi^2(1)$	51.1	29.4	37.1	33.0	2.8	52.6	14.6	40.3
p	< 0.001	< 0.001	< 0.001	< 0.001	0.093	< 0.001	< 0.001	< 0.001

- Stringent Test: Does it predict development? Does it subsume SSP effect?
- Precpecified UG constraints
  - No correlation with SSP
  - Does not derive/subsume developmental SSP effect
- Inducing constraints
  - Correlation with SSP
  - Does not derive/subsume developmental SSP effect
- Similar results for variants

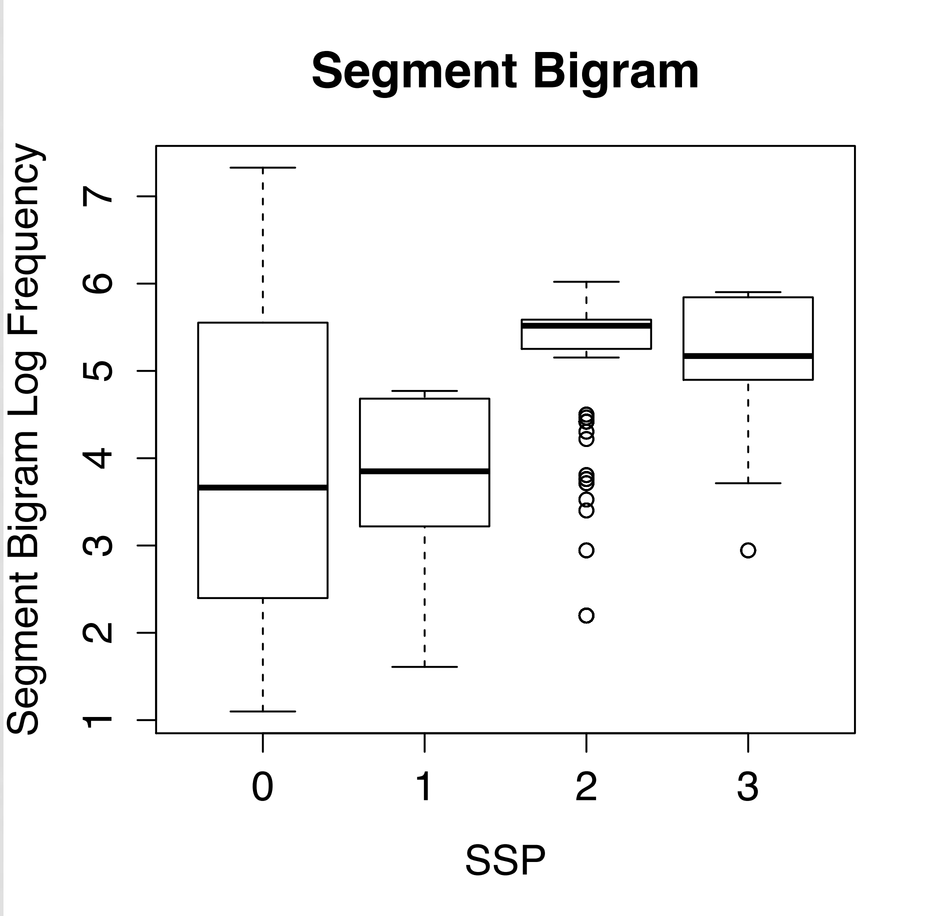
MORE STATISTICS: SSP STILL PREDICTIVE

		Segmental Bigram		Sonority Profile		Sonority Rise	
	Base	+Freq	+Freq+SSP	+Freq	+Freq+SSP	+Freq	+Freq+SSP
D_{xy}	0.425	0.445	0.47	0.437	0.459	0.453	0.458
LR	357.4	392.2	427.0	375.77	412.30	388.65	409.83
$\chi^2(1)$		34.8	34.8	18.4	36.5	31.3	21.2
p		< 0.0001	< 0.0001	< 0.0001	< 0.0001	< 0.0001	< 0.0001

- Other lexical statistics?
  - Phoneme bigram, sonority profile freq., rise frequency?
  - Type, token?
- Results
  - Nope! None subsume SSP effect

EXAMPLE DISTRIBUTION

- Why not?
  - Correlation but
  - Not perfect prediction
- Correlation is not enough!



MODELING SUMMARY

- No lexicalist model derives SSP effect
  - UCLA Phonotactic (inducing constraints)
  - UCLA Phonotactic (fixed constraints)
  - Other lexical statistics
  - Type/token
  - Using finer Sonority scale
- Conclusion
  - Lexicalist models cannot be the full story

PUSHING LEARNING MODELS

- Lexicalist learning models need to be put to harder tests
  - Not just restrictive languages that mirror universals
  - Not just adult judgments, but development too
  - It's not sufficient to correlate with behavioral data
    - Models must predict behavioral data
    - Models must derive universals
- Learning Models must be comprehensibly evaluated on all cases where we see evidence of universals
  - Preferences predicted for acquisition of licit structures
  - Preferences predicted among novel licit structures
  - Preferences predicted among novel illicit structures
  - Ultimately, on predictions for typology

IN PROGRESS

- Joint work with Amanda Rysling
- Experiment with adult native Polish speakers
  - Examine native speaker judgments
    - novel licit structures
    - novel illicit structures
  - Does SSP bias persist?
  - Does it affect licit/illicit structures equally?
  - Test models' predicted preferences
    - novel licit structures
    - novel illicit structures
- Link for potential participants
  - **http://spellout.net/ibexexps/polson/pol-SSP/experiment.html**

GENERAL SUMMARY

- English: Sonority Projection Effects
  - Raw statistics are not enough
  - But the right models can induce effects from input
  - Not a clear case of Poverty of the Stimulus for SSP
- Polish: Plenty of (conflicting) Stimulus
  - SSP is predictive of production accuracy
  - No input-based model predicts SSP effect
- Conclusions
  - Support for 'defiance' of the stimulus in Polish
  - Input sensitivity alone cannot explain development
  - Support for some kind of SSP bias in acquisition
    - Ongoing work to further understand nature of this bias