Apr. 13, 2018 University of Calgary

Phonological regularity, perceptual biases, and the role of phonotactics in speech error analysis

John Alderete, Simon Fraser University

in collaboration with:

Monica Davies (UBC) and Paul Tupper (SFU)

Speech errors tend to respect grammar

"First law" of tongue slips (due to Wells 1951)
Speech errors respect phonotactic constraints, only produce legal phonological combinations (Boomer & Laver 1968, Nooteboom 1967, Garrett 1980)

Phonological constraints active in repairs (Fromkin 1971: 41)
play the victor → *flay the pictor* (exchange of *p* and *v*, *vl* → *fl*)

Syntactic regularity in speech errors (Garrett 1980, Bock 2011)

- Category constraint (word substitutions respect part of speech labels), producing licit but unintended sentences.
- Sentence blends, role mis-assignments, and spurious agreement relations tend to respect grammar too.

Bock 2011: 332
"The most striking thing about attested syntactic errors is that, like other kinds of speech errors, they occur within a structural matrix that resists modification."

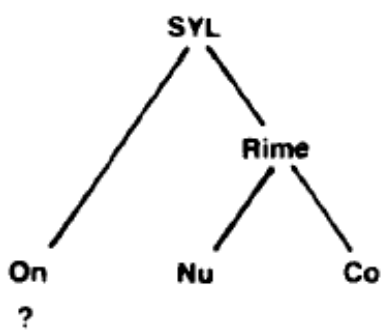
Regularity as a hard constraint

Early models: grammatical regularity not really explained, but the result of a built-in "structural matrix"

Spreading-interactive model of Dell 1986

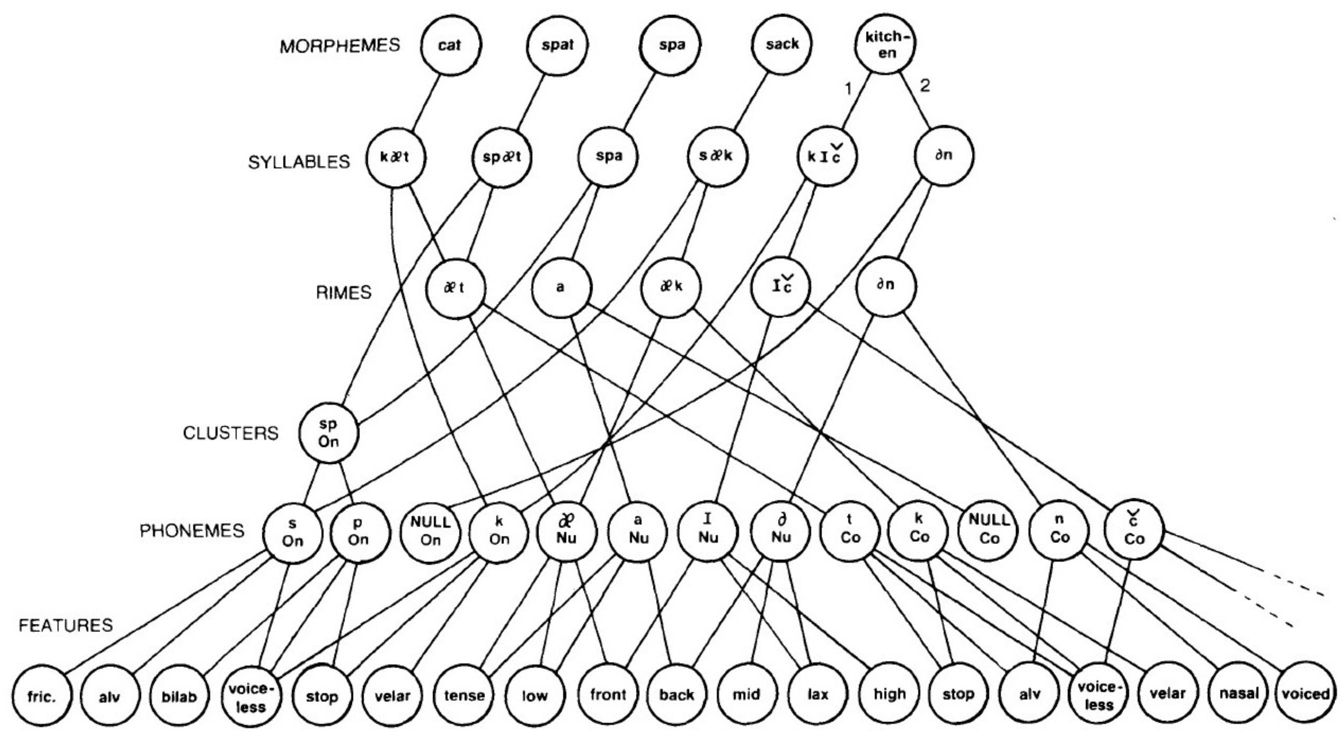
- Mental lexicon: activation dynamics for selecting valid linguistic units
- Tactic frames: productive capacity for language (simple sentence trees, word trees, and syllables. (a.k.a. 'structural matrix') ***This is a clear role for grammar!***

Tactic frame



Result: *vl* is not a valid onset for intended word *play* because [*vl*, Onset] is not a node in the mental lexicon.

Mental lexicon (fragment)



But regularity is not a hard constraint

Stemberger 1983: phonological regularity is very high, but speech errors do violate English phonotactics, approx. 1% of the time (37 violations/6,300 examples); standard of 99% phonological regularity

… in the first floor /**dl**orm —- dorm room

I /**sth**ough—- thought I said 'moff'

… knowledge of the cooperative /**rp**in—- principle

Problem: need a model that can produce phonotactic violations (rarely); not possible if regularity is a hard constraint.

Issue (Stemberger): the dominance of phonological regularity in speech errors does not entail that speech errors are controlled directly by phonotactic constraints—other independently need mechanisms sensitive to frequency could be at work.

Dell et al. 1993: regularity without tactic frames

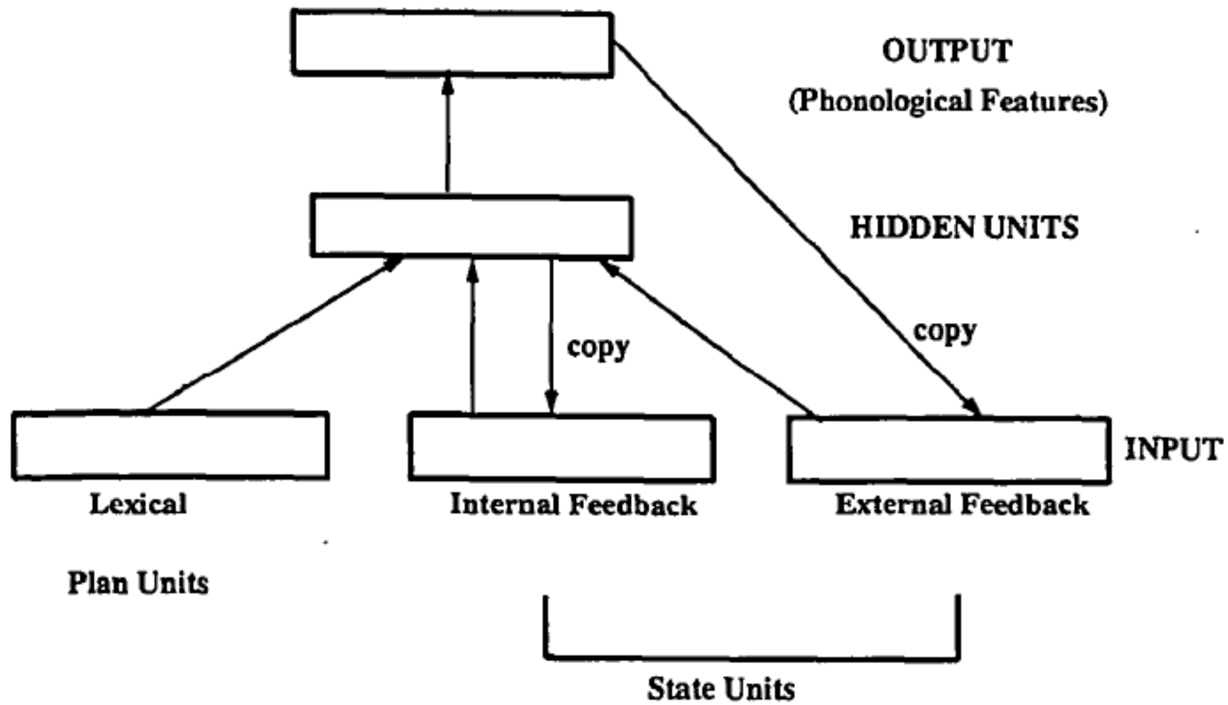
Dell et al. (1993): simple recurrent network proposed as a model of phonological encoding. Trained on a sample of English words and tested against a set of phonological benchmarks characteristic of speech error patterns (i.e., phonological regularity, CV substitutions, syllabic constituent effect, word-onset asymmetry)

Network features:

Sequential: outputs a single segment, then another, in sequence

Recurrent: current segment processed in tandem with knowledge of past segments

Distributed representations: segments are represented as a vector of feature values (cf. distinctive features)



Result: given certain parameters (trained on frequent vocabulary, internal and external input), the model produces errors that are **phonotactically regular about 96.3% of the time**

Upshot: regularity seems to be achievable without tactic frames (But a little below Stemberger's standard of 99%)

Questions

Just how much phonological grammar is there in phonological encoding?
The success of Dell et al's SRN suggests that some phonological structures (e.g., syllable templates) may not need to be formal mechanisms in model of language production processes.

Is cross-linguistic markedness a factor in the structure of speech errors, and if so, how is it incorporate into phonological encoding?

- Markedness is an important ingredient to contemporary phonological grammar (Constraints and Repairs, Optimality Theory, HG, MaxEnt)
- Markedness has also be argued to be a factor in the structure of speech errors (Blumstein 1973, Goldrick 2002, Goldbrick and Rapp 2007, Romani and Calabrese 1998), or not (e.g., Shattuck-Hufnagel and Klatt 1979)
- If markedness is a factor in the structure of speech errors, how is it included in model implementations?

How does methodology affect description of phonological regularity?

- Perceptual biases may reduce the rate of phonologically illicit errors (Dell et al. 1993) and other phonological effects (Alderete & Davies 2018)
- If speech errors are collected in such a way that these biases are reduced, does that effect phonological regularity and other phonological factors?

Focus and approach

SFU Speech Error Database: principal dataset

Describe a methodology for collecting speech errors and demonstrate that it is more reliable and robust to perceptual bias than prior work

Re-assess phonological regularity in English speech errors Just how phonologically regular are speech errors? (How common are phonotactic violations?)

Examine a host of dimensions of cross-linguistic markedness to see if it has a major impact on the structure of speech errors

SFU Speech Error Database (SFUSED)

Current languages

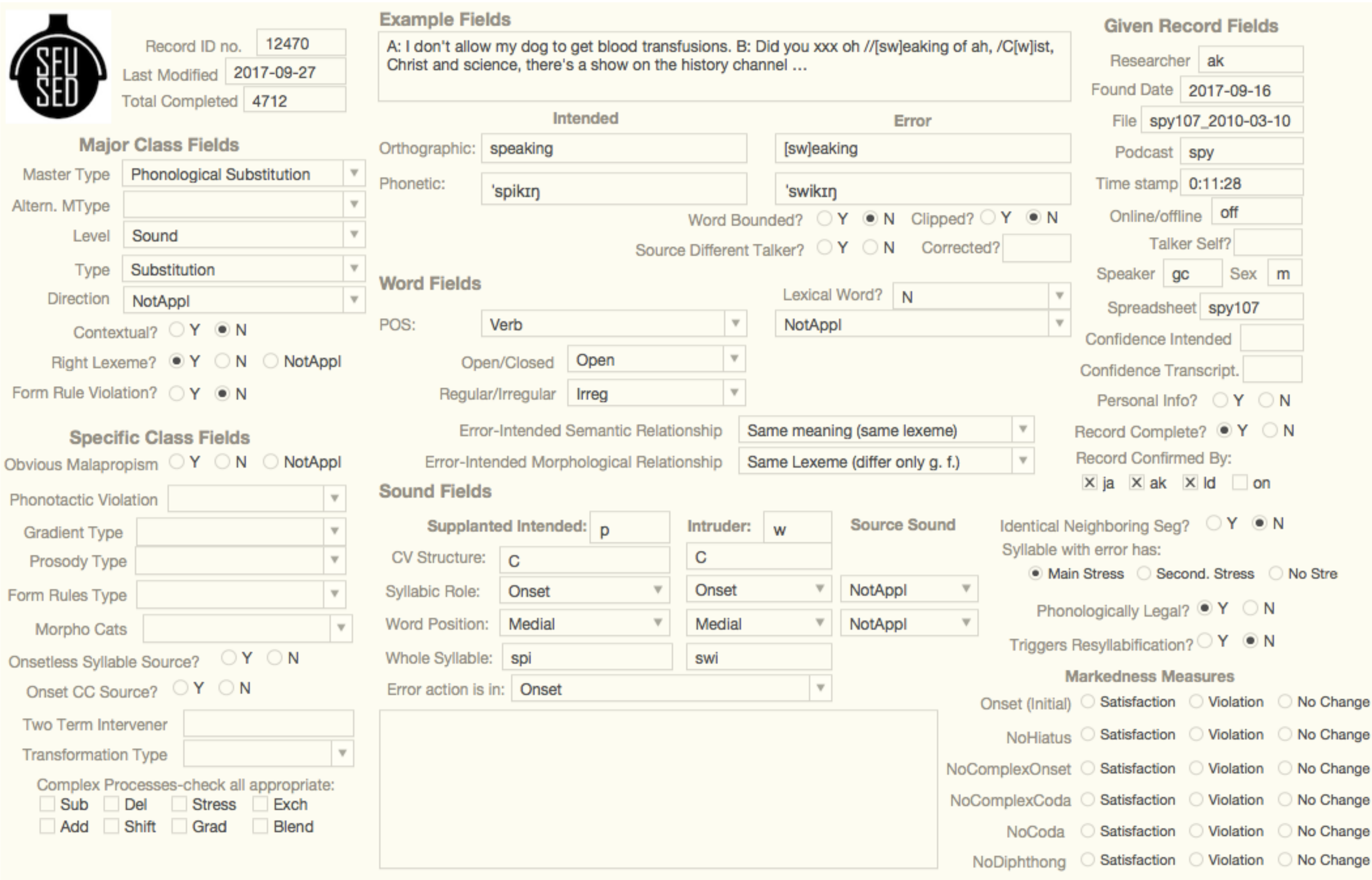
SFUSED English (10,104 errors)

SFUSED Cantonese (2,549 errors)

Goals

- Build a multi-purpose database that supports all types of language production research
- Linguistically sophisticated coding of errors
- Explicit processing assumptions and coding that support probing of psycho-linguistic biases
- Examine how the structure of non-Indo-European languages impacts language production processes

SFUSED English interface



Methods: data from audio recordings

Natural conversations in audio recordings
-errors collected from unscripted speech in podcast episodes
-scripted speech ignored, not counted in time metrics
-viewed as waveform in speech analysis software
-"offline" data collection: can replay and slow down
-cf. "online" data collection in most prior studies: direct observation

Podcasts
-mix of genres: lifestyle, entertainment, science, technology
-selection criteria: high production quality, long segments of unscripted speech, talkers no media professionals, balance of male and female speakers
-nine podcasts, roughly 50 hours of each, about 1000 errors

Dialects and idiolectal features
-"Standard American English", i.e., some variety of Midlands dialect
-idiolectal features: e.g., intrusive velar [mʌktʃ] 'much'; habitual so not an error
-podcast notes: non-standard dialects, idiolectal features, used as reference material for assessing errors

Definition of speech error (Dell 1986)
"unintended non-habitual deviation from the speech plan"

Methods: data collectors and training

Multiple data collectors
-total of 13 data collectors (currently)
-data collector part of the record (can be examined for skewed distributions)
-cf.: 1 or 2 experts studies, or large number of untrained data collectors

Training
-undergraduate students with introduction to formal linguistics
-given phonetic training in transcription and tested for transcription accuracy
-introduction to speech errors, definition and illustration of all types
-training through listening tests: assigned pre-screened recordings, asked to find errors; given feedback

Target for training
-MPE (minutes per error) of 3.0 or less
-less that 25% false positives

Methods: verification and classification

Verification

-data submissions batch imported, verified by data analyst
-checked against reference material on casual speech patterns
-checked for slips of ear, change of speech plan, false starts, errors of ignorance
-podcast notes checked for idiolectal features (habitual, not an error)
-approximate 25% of submitted errors are removed, not true errors

Classification separate from data collection

-data analyst classifies each error with variables from standard taxonomies
-major class: unit, type of operation (e.g., substitution, deletion, etc.), direction
-ambiguity resolved with Occam's Razor (though alternatives retained as part of record)

Advantages

Alderete & Davies 2018, Language and Speech

Reliability and data quality
-audio recording supports data collection separate from verification by another researcher
-with different collectors, can minimize collector bias and measure it if it exists
-audio recordings help in spotting idiolectal features and phonetic structures

Metrics
-audio recordings have a duration, with allows measures that are not possible with online collection, e.g., collection metrics
-supports much better estimates of speech error frequency
-using capture-recapture methods, we find that speech errors are much more frequent than reported in prior work (an error at least every 48.5 seconds, probably more)

Data discovery
-audio recordings allow acoustic analysis, probe fine-grained phonetic detail
-can address frequent cry for “more context” (can be recovered)
-with a time metric, can investigate time-based effects like speech rate

Better sample of true population of speech errors
-sample has much higher coverage, likely three to four times better
-less ‘easy to hear’ and more ‘hard to hear’ speech errors
-collect more errors that occur in fast speech

Sample coverage: less 'easy to hear' errors

Online collection: direct observation, requires high degree of confidence

Offline collection: from audio recordings, can replay and listen to slow speech

Exchanges in SFUSED

Ex. *We can just wrap mine in a* ***/torn /korkilla*** (*corn tortilla*, 1495)

Early data collection had 1,100 errors collected online. Balanced for experience.

	Offline	Online
Morphemes		6
Phrases		1
Sounds	1	25
Words	1	15
Totals	2 (0.38% of 533)	47 (5.6% of 839)

% Exchanges elsewhere:

Stemberger 1982/85: ~6%

Pérez et al. 2007: 35%

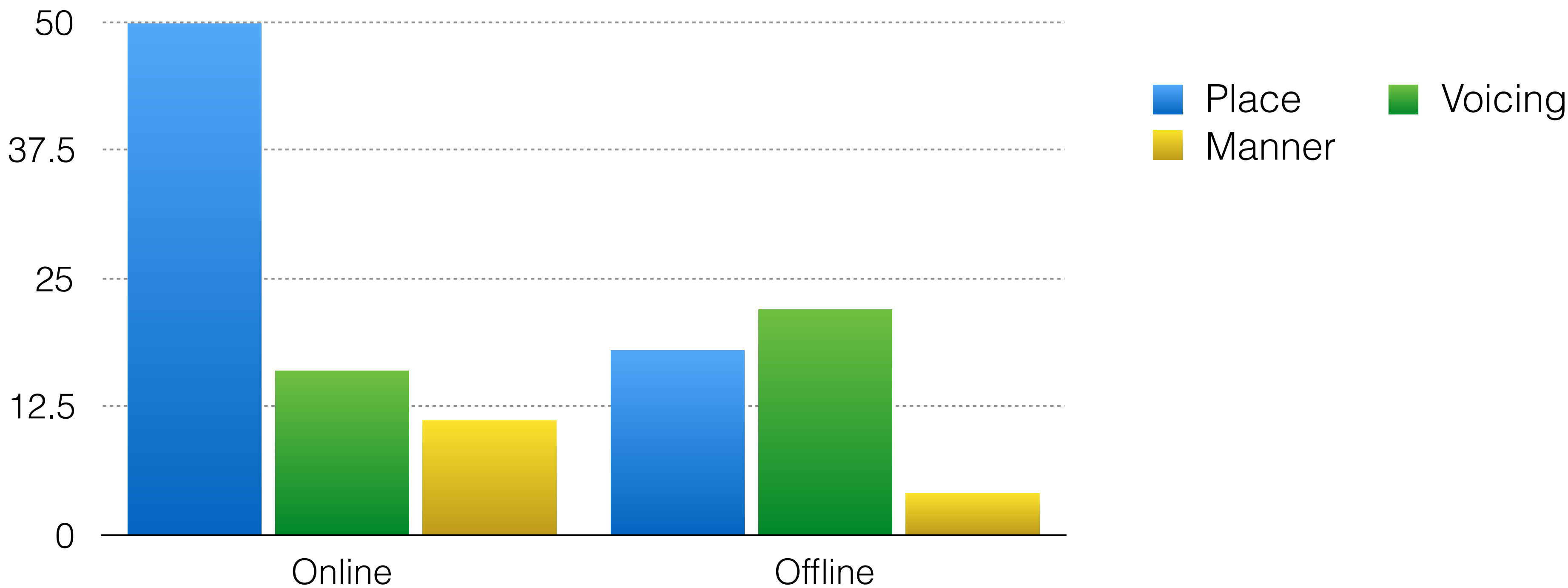
Dell and Reich 1981: 54%

Sample coverage: more 'hard to hear' errors

Finding: errors in mis-pronunciation are easier to detect in place of articulation than voicing (Cole et al. 1978).

Online: counts reflect this perceptual bias

Offline: counts don't reflect this bias; more voicing errors detected than place.



Summary of differences

Alderete & Davies 2018 Language and Speech

Sound errors

- Online errors have more corrected errors than offline errors.
- Online has a stronger repeated phoneme effect than offline errors.*
- Online errors have a stronger lexical bias than offline errors.(*)
- Online errors have a weaker word-onset effect than offline errors.*
- Online errors are more likely to be contextual than offline errors.*
- Online errors have more perseverations and exchanges than offline errors.*
- Online sound substitutions are more symmetric and more concentrated in a small number of substitutions than offline errors, which are more diffuse and asymmetrical.*

Word errors

- Online errors have less additions and deletions and more blends than offline errors.*
- Online word substitutions are much more likely to be in nouns than offline errors, which are more diffuse across lexical and function categories.*
- Online errors tend to respect the category constraint more than offline errors.

* = significant association from chi square test

How does methodology affect data composition?

How phonologically regular are speech errors in SFUSED English?

English phonotactics

Guiding assumption: a word is phonotactically licit if it can be syllabified within a well-formed syllable of English (Kahn 1976, Giegerich 1993, Jensen 1993)

Onset	**Peak**		**Coda**
(s)(C1)(C2)	X4	(X5)	(C6)(C7)(C8)(C9)

Conditions:
All C positions are optional.
Banned C1: *ŋ ʒ* , Banned Codas: *h, j, w.*
Onset clusters: obstruent + sonorant
Appendix + C, C always a voiceless stop, *sf* rare/loans
Banned onset clusters: vd fric/affricate + sonorant, labial + *w*, coronal nonstrident + *l, θw ʃjV ʃw ʃl sr sh gw stw skl*
Onglide *j*: part of peak because of limited distribution, but cannot occur in *CCju* cluster.
Coda clusters X5+C6: falling sonority (r > l > nasals > obstruents) and *s* + *p t k; lg* is banned.
C7-9 are appendices limited to coronal obstruents
Nasal + obstruent clusters agree in place and the obstruent is voiceless.
Tense vowels and diphthongs are bimoraic (fill X4 and X5), lax vowels are short fill X4.
Stressed and final syllables are bimoraic (lax vowels occur in closed syllables) and all syllables maximally trimoraic (syllables tense vowels only have simple codas)

Results: illustrations of phonotactic violations

Alderete & Tupper, to appear
WIREs Cognitive Science

Substitutions
Illicit onsets/appendices
1500 … by the maps at the ^selection /[**ʃkr**in] (screen)
5739 … they shoot, /[**ʒ**u] shoot The Thick of It … (you)

Illicit codas/rimes
1245 … Their HOV /[l**aɪŋ**] xxx lane is like one driver (lane)
5898 Vin Diesel got kicked off of /Rei[**ŋ**]deer Games … (reindeer)

Nonnative sounds
5964 … first of all, Katrina /[kl**y**]= clearly defined (clearly)

Additions
Illicit onsets, appendix + onset
49 … get the Ferrari down a /[**flj**u] xxx few ^floors? (few)
1278 I don't like the ^/**vr**iral ^marketing. (viral)
5599 … talking a ^dream, what that ^dream /[**mr**]eans … (means)

Illicit codas/rimes
1526 The ^person /[k**eɪmp**] ^up to the desk.

(SFUSED record ID # on left)

Results: illustrations, cont'd

Deletions
3954 … Lisa, /**Sr**eech and Lisa. (Screech)
8943 … I think you're a /h**u[ŋə**]= hunk-a-rama.

Exchanges
4581 ... the children in the trailer for /Moon[raɪ**ŋ**] /Kee**z**= Moonrise Kingdom.

Sequential Blends
4453 … A diary is a /[**sb**]ook xxx a very special book.
5278 … you can't quite /[**pj**ɪrt] xxx put your finger on.
7211 … because we /[spɪ**lkf**] xxx we, we speak film

Word Blends
870 … ... /[sɑs**tæ**] makes me frisky. (pasta, sauce)
7120 Top ten /**thw**ays to make me cry (things, ways)
7270 … /so[m-**bw**ʌn-di] xxx uh in the … (someone, somebody)

Results by error type

Observations: % of phonotactic violation differs by type, but overall % of irregularity much higher than 1% found in Stemberger's corpus.

Error type	Example	*N*	Violations	% of *N*
Substitutions	pleep for *sleep*	1,376	44	3.20
Additions	bluy for *buy*	358	33	9.22
Deletions	pay *for play*	169	3	1.78
Exchanges	heft lemisphere for *left hemisphere*	37	2	5.41
Shifts	*splare backforests* for *spare blackforests*	7	0	0.0
Sequential Blends	Tennedy for *Ted Kennedy*	57	4	7.02
Word Blends	tab for *taxi/cab*	72	4	5.56
Totals		2,076	90	**4.34**

Perceptual bias: missed phonotactic violations

Conjecture: Cutler 1982, Shattuck Hufnagel 1983 point out that there is probably a perceptual bias against phonotactic violations. Listeners may regularize them or simply fail to hear them.

Probe: Alderete and Davis 2018 used balanced sample of online vs. offline errors and found a significant association between methodology and regularity ($\chi(1)2=7.902$, $P=0.0049$).

	Offline	Online
Phonotactic Violations	17 (3.19%)	8 (0.95%)
No Violations	516 (96.81%)	831 (99.05%)

Perceptual bias: all sound errors

Conjecture: Cutler 1982, Shattuck Hufnagel 1983 point out that there is probably a perceptual bias against phonotactic violations. Listeners may regularize them or simply fail to hear them.

Probe: counting all sound errors and blends, % of phonotactic violations higher ($X2 = 16.9618$, $p < .05$); note effect does not depend on what counts as a violation.

	Offline	**Online**
Phonotactic Violations	76 (5.5%)	11 (1.6%)
No Violations	1,326 (94.5%)	660 (98.4%)

Overwhelmingly regular, but above chance?

Question: the lower standard of phonological regularity raises the question of whether significantly above chance levels; intended words are regular, so perhaps just commitment to speech plan is enough.

Permutation test (see Dell & Reich 1981)

1. Randomly permute segments from a list of intruder segments (given from error corpus), holding constant the phonological context (i.e., segments selected independent of context)
2. Use multiple trails to obtain a distribution of the percentage of regular errors under the independence assumption.
3. Test to see if there is sufficient evidence to reject independence hypothesis.

Example: C1 slot in substitutions

1. Take all intruders in C1 slot of <u>C1</u>C2 complex onsets in the error data, randomly permute them in the same slots they occur in errors.
2. Permute segments of C1 randomly many times, get an average rate of phonologically regular words
3. Compare with actual errors in C1 of C1C2 errors to see if deviate significantly.

Results: some, not all, above chance

Finding: in both substitution and addition errors, significantly above chance in non-initial positions (C2 of cluster), but not initially. No such evidence for initial errors.

Type	Context	Example	*N*	Actual	Random	Significant?
Substitutions	_C of $\underline{C}C_{\text{Onset}}$	*blue* → *plue*	37	81%	78%	No (p=0.38)
	C_ of $C\underline{C}_{\text{Onset}}$	*dream* → *dweam*	36	100%	83%	Yes (p=1e-6)
Additions	_C into $\underline{C}C_{\text{Onset}}$	*last* → *flast*	29	62%	64%	No (p=0.77)
	C_ into $C\underline{C}_{\text{Onset}}$	*bad* → *brad*	75	87%	79%	Yes (p=0.005)

Interpretation:

- Non-initial contexts require analysis; 'structural matrix' of some kind to predict above chance level.
- C1 errors are dominated by errors that occur word-initially, so could be an effect of the word-onset bias (Wilshire 1999)

Discussion: comparison with SRN

New standard: 93-95% phonologically regular, cf. 99% of Stemberger 1983, (94.5% regularity reported in offline data still probably affected by perceptual bias)

Goodness of fit: Dell et al. 1993 simple recurrent network tested a variety of parameters that compare well with these findings.

- Models trained on frequent vocabulary and with both internal and external representations: **96.3% regularity**
- Range for other assumptions about input: **89-95% regularity**
- Many of the errors with phonotactic violations resemble the phonotactically illicit errors we have found, with illegal clusters and initials.
- Model also predicts more errors word-initially, so accounts for contextual differences in permutation test

Limitations

- Model trained only on three segment words, so no polysyllabic words
- Didn't really allow for additions, which account for a lot irregularity (perhaps 1/3)
- Phonotactics likely slightly different than one used here (likely less stringent).
- Didn't account for prosody (stress in errors) and other structures.

Take home: with the new standard, tactic frames (cf. syllable templates) are not obviously necessary to the analysis of phonotactic regularity in speech errors.

Discussion: other potential roles of grammar?

Markedness

- Markedness is “the stuff” of most grammars in contemporary phonology. If we could find a role for markedness, this would be a clear role.
- Focus is on cross-linguistic markedness, not language particular markedness relations, because latter is hard to separate from frequency.

Frequency

- Frequency structure is increasingly a part of formal grammar, e.g., weights in Harmonic Grammar and MaxEnt grammar.
- Often overlaps markedness (marked is less frequent), but not always.
- Language production: frequency is a standardly assumed output bias in language production (Dell 1986)

Feature specification (‘anti-frequency’)

- Feature specification is a core assumption in linguistic grammar, and has also been argued to account for speech error facts (Stemberger 1991)
- Specified sounds (because contrastive) override unspecified (because predictable): e.g., palatal bias in consonant substitutions.

How does segmental markedness shape speech errors?

Background: segmental markedness

Bias for marked → unmarked mappings in speech errors at segmental level

Experimentally induced speech errors

- Kupin 1982: disyllabic tongue twisters, unmarked forms preferred
- Goldrick 2002: implicit learning paradigm, examined substitutions where markedness and frequency make different predictions.

  Example: [t] is unmarked relative to [s], but less frequent

  [s] → [t] > [t] → [s] supports markedness account

Aphasic speech

- Blumenstein 1973: single feature consonant substitutions favour marked → unmarked mappings (just Broca's and Wernicke's aphasics, not conduction aphasics)
- Romani et al. 2002: markedness superior to frequency in aphasic consonant substitutions
- Goldrick and Rapp 2007: brain-damaged subject with deficit in post-lexical phonological processes, more accurate with coronals /t d/ (93%) than dorsals /k g/ (86%)

Against markedness as a factor

Some studies have found no effect of markedness, and segment substitutions reflect baseline frequencies ('availability'): Shattuck-Hufnagel & Klatt 1979, Stemberger 1991

Focus: single consonant substitutions

Example: ... we ^read out what's in each pixel
/**r**un ^row at a time. (Intended: *one*, *w* → *r)*

- Most common type of error in all speech error corpora in normal speakers, so able to get sufficient data
- Can study segmental markedness by considering differential effect of markedness on feature structure.
- Cross-linguistic markedness makes predictions about marked and unmarked values of features: voiced marked relative to unvoiced
- Prior (and current) research: consonant pairs that differ in a single feature.

Test: single feature consonant substitutions

Procedure: take a consonant confusion matrix (*N= 1,506)*

Sound Intruder

Sound Suppl..	p	b	t	d	k	g	f	v	θ	ð	s	z	ʃ	ʒ	tʃ	dʒ	m	n	ŋ	l	r	w	j	h
p		27	10		10		10	1							2		2	1						
b	14	1	1	9	1	1	1	8			1						7			3		2		
t	7	2		15	13	1	2	1	2	2	2	3	3		10			7		6	1		1	1
d	1	8	3		1	6			2	2	3	3				6		5		2	3			
k	10	2	17	2		16					2	1	2		3	1			1	1	1			2
g		3		2	15		1	1			1		1					1						
f	8	3	2			1		6	3		9	1								1	1			1
v		9	1		1	1	11			2		2				2	1	1		1	3	3		
θ	2		3	2	2		1			2	13	1	1		1					1				
ð		1	3	3		1		2	2		1	1			1	1	1			4				1
s		1	8	1	4	1	4	1	12	1		7	55		5			2		1				1
z		1		2		1		2	3	1	8		4	1		2	1		1	2				
ʃ				1		1	1		1		14	1		1	6						1			
ʒ											1													
tʃ			4		4						5		5	3		1					1		1	
dʒ			1	7	1	3					1			1	5							1		
m	2	8					1	1		1		1						24		3		6		
n			3	12	1						1	1					16		4	8	5		1	1
ŋ																		1						
l		1			1	3				2			1				4	3	2		21	6	2	
r	1			1				1		1					1	1	3			41		12	2	
w	1	2	1	2			1	1									4			8	15			
j		1				2								1		1				2	3			
h		1	4	1	4	3	1		1		3				1					2				

Test: single feature consonant substitutions

Procedure: take a consonant confusion matrix (*N= 1,506)*

	Sound Intruder																							
Sound Suppl..	p	b	t	d	k	g	f	v	θ	ð	s	z	ʃ	ʒ	tʃ	dʒ	m	n	ŋ	l	r	w	j	h
p		27	10		10		10	1							2		2	1						
b	14	1	1	9	1	1	1	8			1						7			3		2		
t	7	2		15	13	1	2	1	2	2	2	3	3		10			7		6	1		1	1
d	1	8	3		1	6			2	2	3	3				6		5		2	3			
k	10	2	17	2		16					2	1	2		3	1			1	1	1			2
g		3		2	15		1	1			1		1					1						
f	8	3	2			1		6	3		9	1								1	1			1
v		9	1		1	1	11			2		2				2	1	1		1	3	3		
θ	2		3	2	2		1			2	13	1	1		1					1				
ð		1	3	3		1		2	2		1	1			1	1	1			4				1
s		1	8	1	4	1	4																	
z		1		2		1																		
ʃ				1		1	1																	
ʒ																								
tʃ			4		4																			
dʒ			1	7	1	3																		
m	2	8					1																	
n			3	12	1																			
ŋ																								
l		1			1	3																		
r	1			1																				
w	1	2	1	2			1																	
j		1				2																		
h		1	4	1	4	3	1																	

Test: examine consonant pairs that differ in a single feature, adjust for baseline frequencies

Mapping	Count	Baseline
p → b	27	*p* produce 50 times in 1000
b → p	14	*b* produced 29 times in 1000

Baseline frequencies: estimating relative risk

	Event	General Population
Condition 1	a	b
Condition 2	c	d

mutually exclusive

$$RR = \frac{a/b}{c/d}.$$

see Stemberger 2007

Baseline frequencies: estimating relative risk

	Event	General Population
Condition 1	a	b
Condition 2	c	d

mutually exclusive

$$RR = \frac{a/b}{c/d}.$$

Null hypothesis: RR = 1

	Voicing [s z]	Token Frequency
z → s	8	16
s → z	7	99

$RR(\mathrm{sz}) = 7.07$

Baseline frequencies: estimating relative risk

	Event	General Population
Condition 1	a	b
Condition 2	c	d

mutually exclusive

$$RR = \frac{a/b}{c/d}.$$

	Voicing [s z]	Token Frequency
z → s	8	16
s → z	7	99

$$RR(\mathrm{sz}) = 7.07$$

Test results: are the observed differences significant (not due to chance)? And if so, what direction (favour marked or unmarked structure?)

Baseline frequencies: estimating relative risk

	Event	General Population
Condition 1	a	b
Condition 2	c	d

mutually exclusive

$$RR = \frac{a/b}{c/d}.$$

	Voicing [s z]	Token Frequency
z → s	8	16
s → z	7	99

RR(sz) = 7.07

Test results: are the observed differences in probability of two events significant (not due to chance)? And if so, what direction (favour marked or unmarked structure?)

95% confidence interval (testing null hypothesis that log(RR) = 0):

$$\log RR \in \left(\log\frac{a/b}{c/d} - 1.96\sqrt{\frac{1}{a}+\frac{1}{c}}, \log\frac{a/b}{c/d} + 1.96\sqrt{\frac{1}{a}+\frac{1}{c}}\right)$$

Agresti 1996

Example: log(*RR*)= 1.956, 95% confident log(*RR*) =/ zero, can reject null hypothesis. Direction (sign): favours unmarked segment [s].

Results: [voice], [anterior], [continuant], [nasal]

Voicing

Unmarked	Marked	Direction	Significant?
p	b	marked	N
t	d	marked	N
k	g	unmarked	N
f	v	unmarked	N
s	z	unmarked	Y

Continuancy

Unmarked	Marked	Direction	Significant?
p	f	unmarked	N
b	v	unmarked	N
t	s	unmarked	Y
d	s	unmarked	N

Anteriority

Unmarked	Marked	Direction	Significant?
s	ʃ	unmarked	N
t	tʃ	unmarked	Y
d	dʒ	unmarked	Y

Nasality

Unmarked	Marked	Direction	Significant?
b	m	marked	N
d	n	unmarked	N

Finding: 4 of 14 consonant pairs reached 95% significance, all in the direction predicted by markedness (some pairs not reported due to insufficient data)

Results: place features

Coronal - Labial

Unmarked	Marked	Direction	Significant?
t	p	unmarked	Y
d	b	unmarked	Y
n	m	unmarked	Y
s	f	unmarked	Y

Labial - Dorsal

Unmarked	Marked	Direction	Significant?
p	k	marked	N
b	g	unmarked	N

Coronal - Dorsal

Unmarked	Marked	Direction	Significant?
d	g	marked	N
t	k	unmarked	Y

Finding: majority of place-changing substitutions significant, especially those involving coronals.

Markedness distinct from frequency bias?

Feature	Unmarked	Marked	Direction	Significant?	Frequency bias?
[voice]	s	z	unmarked	Y	Y
[anterior]	t	tʃ	unmarked	Y	Y
[anterior’	d	dʒ	unmarked	Y	Y
[continuant]	t	s	unmarked	Y	N
Place	t	p	unmarked	Y	Y
Place	d	b	unmarked	Y	Y
Place	n	m	unmarked	Y	Y
Place	s	f	unmarked	Y	Y
Place	t	k	unmarked	Y	Y

Confound: while there are many significant results supporting a role for markedness, 8 of the 9 cases could be explained with an output bias for frequent segments (type frequency, interactivity in the lexicon). [t] → [s] is the same mapping Goldrick (2002) found to support the markedness account using experimental methods.

Take home: weak support for a role for markedness in consonant substitutions.

More potential markedness effects

Coda condition effects: no clear effect of markedness

Step sample of consonant substitutions found no preference for replacing non-coronals with coronals in coda position (both medial and final codas).

- Non-coronal → non-coronal > non-coronal → coronal
- No distinction between coronal → coronal and coronal → non-coronal
- Same patterns found in non-contextual substitutions only.

Ban on tense vowels in CVCC syllables: baselines rather small, hard to tell

- Expectation: if speech errors are sensitive to this constraint, expect additions that create a CVCC syllable to prefer nuclei with lax vowels
- Finding: small sample (total 14 observations), but roughly half additions have lax vowels and half have tense vowels

Syllable structure constraints in deletions and additions: unclear

- Marked syllable structures: #_V, V_V, CVC, CCV, CVCC
- Large numbers of examples, especially with phonological addition errors.
- Deletions: show strong effect of syllable structure constraints, i.e., typically reduced marked consonant clusters, and onsetless syllables
- Additions: equal and opposite effect; very often lead to violations of markedness constraints; perhaps due to nature of phonotactics and general addition bias in errors

Summary: cross-linguistic markedness

Segmental markedness

- 9 of 29 consonant pairs that differ in one feature show an effect of markedness
- But only 1 (t>s) of these 9 is best explained as a preference for unmarked segments.

Syllable structure effects (preliminary results)

Marked syllable structures: #_V, V_V, CVC, CCV, CVCC

- Deletions seem to favour marked → unmarked
- But additions favour the opposite order, unmarked → marked to the same degree
- Consonant substitutions don't seem to favour independent Place specification in codas (Coda Condition effects)
- Consonant additions don't seem sensitive to tense/lax distinction in CVCC syllables

Take homes

New standard for regularity: errors of phonological encoding are phonotactically regular about 94-95% of the time (cf. prior standard of 99%).

Methodology: this new standard is documented using methods less prone to bias; so lower standard is likely the result of perceptual bias.

Grammar in phonological regularity? Syllable structure templates do not seem to be necessary to overall phonological regularity. Models of phonological encoding without them seem consistent with the fact.

Cross-linguistic markedness? A detailed examination of segmental (and syllable structure) markedness did not reveal a strong role for markedness independent of frequency.
But stay tuned …

Implications for speech production models?

Ingredients of a sufficient model

- Output bias for frequent sounds (consonant confusions)
- Output bias for frequent sequences (general phonotactics)

Models that meet these criteria

- SRN of Dell et al. 1993
- Original Dell-net of Dell 1986 (with some overkill)
- Harmonic Grammar with weights for important constraints (Goldrick and Daland 2009)
- Two step interactive model (Dell et al. 1997)

Explanation of frequency effects

- Stems from interactivity in lexical network: type frequency effects a direct result from links to actual words.
- Frequency from learning: repeated exposure to words re-enforces associations between plan and sequence.

Implications: but we still do need syllables …

Missing: explicit syllabification algorithm for organizing segments in a frame.

Roles for syllables in phonological encoding:

Syllable position constraint (Boomer and Laver 1968, Fromkin 1971)
Onsets slip with onsets, codas with codas, etc.

Syllable errors in languages like Mandarin (Chen 2000)
Whole syllables slip at rates greater than expected by chance (probably only in languages with very small syllabaries)

Masked Priming (Ferrand, Segui and Grainger 1996, cf. Schiller 1998)
Shorted masked syllables speed up picture naming if exact syllable in test word.

Implicit priming (Chen, Chen, Dell 2002, cf. Meyer 1991)
Syllables implied in production planning because facilitate picture named in cued recall experiments (may be a language particular effect).

Syllables are stored

Common assumption: Syllables can be stored in the mental lexicon without being actively generated in tactic frames. Assumed in most models built off the Dell (1986) spreading-activation model, and some have argued that languages like Mandarin actually represent and select syllables in phonological encoding (see O'Seaghda et al. 2010)

Syllable position constraint: filler-role bindings from Legendre and Smolensky 2006, syllable position a kind of similarity effect.

Syllable errors and implicit priming in Mandarin: syllables are stored and selected in phonological encoding (see 'proximate unit' hypothesis of O'Seaghdha et al. 2010)

Masked priming: if syllables are stored, then should prime later words that have identical syllables.

Conclusion: while a syllable template does not appear to be intrinsic to phonological encoding, syllables seem to be necessary representations in the mental lexical and production plannng.

Contributors to SFUSED

Director/analyst/data collector:
John Alderete

Research associates
Paul Tupper (SFU)
Alexei Kochetov (Toronto)
Stefan A. Frisch (USF)
Monica Davies (UBC)
Henny Yeung (SFU)

Analysts/data collectors
Holly Wilbee (English)
Monica Davies (English)
Olivia Nickel (English)
Queenie Chan (Cantonese)
Macarius Chan (Cantonese)
Heikal Badrulhisham

Data collectors
Jennifer Williams (English)
Julie Park (English)
Rebecca Cho (English)
Bianca Andreone (English)
Dave Warkentin (English)
Crystal Ng (Cantonese)
Gloria Fan E (Cantonese)
Amanda Klassen (English)
Laura Dand (English)

Why are we still collecting speech errors?

Problem: speech errors ‘in the wild’ are very time-consuming, prone to mistakes in observation and interpretation; often can’t get enough data from a particular pattern to test specific hypothesis.

Stemberger 1992: actually there is considerable overlap in the patterns of errors collected in naturalistic and experimental settings. So speech errors ‘in the wild’ present valid data patterns worthy of analysis.

Some patterns no suitable for experimental study: % of exchanges, lexical bias, non-native segments, phoneme frequency effects, etc.

This research shows that a new approach to data collection (offline, many listeners), has potential for new observations, e.g., phonological regularity

Large databases can be re-purposed and extended, not really true of experiments.

Offline methodology is actually very efficient (see Alderete & Davies 2016 for research costs estimates); can produce a database of 3,000 errors in about the same amount of time it takes to run two experiments.

Idiolectal features are _very important_ in understanding errors (habitual, so not an error), but can only really analyze them after a few hours of listening to a single talker.

Estimating error frequency

Alderete & Davies 2018, Language and Speech

Prior assumption: speech errors are rare in general (**error every 5-6 minutes**), motivates focus on normal language production

Problem: prior estimates of error frequency based on online collection, and many failed to address the fact of missed errors (though all studies concede they miss them).

Capture-recapture: common tool in ecology for estimating a population when exhaustive is impossible or impractical

Take home: speech errors occur much more commonly than enumerated in prior research, at least as often as **48.5 seconds** (upper bound because of non-homogeneity)

Second	A	B	C	AB	AC	BC	ABC	n	$\tilde{m}$	$\tilde{v}$	SPE
2,100	2	18	3	2	0	3	5	33	16.3	49.3	42.60
1,690	6	5	4	5	0	2	9	31	13.48	44.48	38.00
1,993	2	9	5	1	0	1	5	23	20.08	43.08	46.26
2,385	6	6	5	8	2	1	5	33	11.7	44.70	53.36
4,143	24	9	1	5	1	1	3	44	21.84	65.84	62.93
3,000	9	2	7	3	5	1	2	29	10.63	39.63	75.70
1,800	9	9	3	2	0	1	1	25	29.87	54.87	32.81
2,377	15	2	4	3	2	1	3	30	13.39	43.39	54.78
2,400	18	4	6	1	2	0	7	38	41.93	79.93	30.03