June 3, 2017 SFU Cogs Workshop

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

John Alderete, Simon Fraser University

Speech errors

Speech errors are **unintended and non-habitual deviations from a speech plan**.

Sound error: … *what's in each pixel /**r**un **r**ow at a time**.** (one → run*, 27)

Word error: *This side has green Vs, I mean red Vs* (797)

Syntactic error: *I'm not staying all up night.* (1817) (#'s = SFUSED record ID)

Speech errors are common occurrences in human talkers, generally thought to be the **by-product of normal production processes** (not pathological)

Speech errors are patterned, and exhibit a host of psycholinguistic biases, like the lexical bias (sound errors more likely to result in actual word than they would by chance)

Speech errors tell us something about the **nature of language production processes**

Background: Dell & Reich 1981, Dell 1986, Fromkin 1971, 1973, Garrett 1975, Harley 1984, 1996, Shattuck-Hufnagel 1979, Stemberger 1982/1985

Some themes in speech error research

Interactivity of linguistic levels
Examines how the outcomes of specific production processes are affected by the processing at other levels (Dell 1986, Rapp & Goldrick 2000, Stemberger 1985, Vigliocco & Hartsuiker 2002)

Speech errors and serial order
Speech errors reveal a malfunction in the normal processes of ordering chunks of a speech stream (Acheson & MacDonald 2009, Dell et al. 1997, MacKay 1970, Shattuck-Hufnagel 1979)

Structure of speech errors and statistical structure of language
Speech errors are affected by output biases that stem from the statistical structure of language (Dell 1986, Dell et al. 2000, Frisch 1996, cf. Stemberger 1991)

Speech errors and linguistic structure
Probes the psychological reality of linguistic structure by studying the structure of speech errors (Berg 1987, Fromkin 1971, Kubozono 1989, Stemberger 1983)

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 by 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"

Example: Spreading-interactive model: Dell 1986
Mental lexicon: activation dynamics for selecting valid linguistic units

Tactic frames: productive capacity for language (builds sentence trees, word trees, and syllables. (a.k.a. 'structural matrix')

Result: speech errors respect grammar, e.g., syllable canons because of tactic frames and encoding of sounds in the mental lexicon are 'hard-wired' into the system.

Intended word: *play*

Tactic frame for syllables

Onset Peak Coda

_____ _____ _____

Sound selected by activation dynamics

Options for Onset: *pl fl kl gl*

Non-options: *ŋ vl tl bw*

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); supports 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

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 … so still don’t know precise role of grammar here.

Dell, Juliano, Govindjee (1993)
Develop a model of phonological encoding (simple recurrent network) that learns sequences, show that it accounts for phonological regularity (sometimes to lower standards) without tactic frames.

Empirical challenges

Provocative question: Just how regular are speech errors?

Syntactic errors: very rare (approx. 2%), not exhaustively studied
Pilot study in SFUSED

- Majority word shifts are rampantly ungrammatical
  Example: *I'm not staying all up night.* (… *staying up all night*, 1817)
- Many other types (blends, agreement errors) are also frequently ungrammatical
  Example: … *in life I'd never will understand* (348)
- Conjecture: perhaps assumed syntactic regularity is due to a talker's commitments to the message, and not structural requirements.

Phonetic research in speech errors

- Evidence for gradient errors, not simple mis-selection of one phonological category for another.
- Mower & MacKay 1990: phonotactics is not a good way of assessing errors because most sublexical (sound errors) are not errors that make reference to symbolic categories.

A conjecture about perceptual bias (Dell et al. 1993)
It is possible that current assessments of regularity reflect a higher standard, because it is well know that speech error collection is subject to perceptual bias (see Bock 1996); possible that many errors that violate e.g., phonotactics are regularized or simple not detected

Research focus

Just how regular are phonological speech errors? (phonetic errors aside)
→ Quantitative assessment in SFUSED

How does methodology affect phonological regularity?
→ SFUSED designed to be robust to perceptual bias, can investigate Dell et al's conjecture

What is the role, if any, of phonological grammar in explaining the structure of speech errors?
→ Given: phonotactics as hard constraints not valid (many violations)
→ Can we get by without frame constraints? (no explicit phonotactics, just frequency effects, à la Dell et al. 1993)
→ Are there more nuanced effects of phonological grammar, stochastic effects (Goldrick and Daland 2009, Goldbrick 2001)

The approach

SFUSED (= Simon Fraser University Speech Error Database)
Explain the methods used to collect and classify speech errors

Phonotactics
Need an account of English phonotactics in order to know when they are violated.

Regularity and other grammatical effects

- Can assess phonological regularity as a categorical effect (violates phonotactics or not), relates to prior research
- Can also investigate gradient effects of grammar, not an absolute but a trend

SFUSED

Current languages
-English (10,000 errors)
-Cantonese (2,200 errors)

Goals
-Build a multi-purpose database that documents the rich structure in spontaneous speech errors.
-Employ methods that are robust to problems found in other corpora.
-Examine how the structure of non-Indo-European languages impacts language production processes
-Partner with other labs to investigate new problems

Key methodological decisions

Offline collection from audio recordings (see Chen 1999 on Mandarin speech error database)
-errors collected from podcasts on different topics
-podcasts selected for having natural unscripted speech, usually Western Canada and U.S. (Midlands dialect 'Standard American')
-multiple podcasts (8 currently) with different talkers, approx. 50 hours of each podcast
-record dialectal and idiolectal features associated with speakers (because habitual, so not an error); listeners develop expectations about individuals

Multiple data collectors
-reduces collector bias, allows it to be studied (collector ID associated with all records)
-total of 13 data collectors

Training regime
-undergraduate students with introduction to formal linguistics, phonetics and phonology
-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; learn by reviewing correct list of errors. Trainees that reach a certain level of accuracy continue.

Classification separate from data collection
-data collectors use established protocol for finding errors in audio recordings, submit errors in spreadsheet format
-data analysts (must be different than collector) verify the error, classify it using the SFUSED fields

SFUSED Classification

SFU SED

Record ID no. 7879
Last Modified 2017-04-26
Total Completed 4207

Major Class Fields

Master Type: Phonological Substitution
Altern. MType:
Level: Sound
Type: Substitution
Direction: NotAppl
Contextual? ○ Y ◉ N
Right Lexeme? ◉ Y ○ N ○ NotAppl
Form Rule Violation? ○ Y ◉ N

Specific Class Fields

Obvious Malapropism ○ Y ○ N ○ NotAppl
Gradient Type
Prosody Type
Form Rules Type
Morpho Cats
Two Term Intervener
Transformation Type
Complex Processes-check all appropriate:
☐ Sub ☐ Del ☐ Stress ☐ Exch
☐ Add ☐ Shift ☐ Grad ☐ Blend

Example Fields

A: I still use the laser comb. B: Daily. A: /[i]s it the diet or is it the laser comb?

	Intended	Error
Orthographic:	is	[i]s
Phonetic:	ˈɪz	ˈiz

Word Bounded? ○ Y ◉ N Clipped? ○ Y ◉ N
Alt.Level
Corrected?

Word Fields

Lexical Word? N
POS: Verb | NotAppl
Open/Closed: Open
Regular/Irregular: Irreg
Error-Intended Semantic Relationship: Same meaning (same lexeme)
Error-Intended Morphological Relationship: Same Lexeme (differ only g. f.)

Sound Fields

	Supplanted Intended: ɪ	Intruder: i	Source Sound
Syllabic Role:	Nucleus	Nucleus	NotAppl
Word Position:	Initial	Initial	NotAppl

Identical Neighboring Segment? ○ Y ◉ N
Syllable with error has: ◉ Main Stress ○ Second. Stress ○ No Stress

Given Record Fields

Researcher: gf
Found Date: 2015-09-28
File: whc131_2015-05-07
Podcast: whc
Time stamp: 0:06:18
Online/offline: off
Talker Self?
Speaker: ac Sex: m
Spreadsheet: 2015f
Confidence Intended
Confidence Transcript.
Personal Info? ○ Y ○ N
Record Complete? ◉ Y ○ N
Record Confirmed By:
☐ ja ☐ hw ☐ md ☒ on

Methods matter!

Online data collection (most prior research): requires on-the-spot observation, only collect errors with high degree of confidence.

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

Result: speech error patterns in SFUSED more diffuse, less concentrated in highly salient errors like blends or exchanges.

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

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

From *Alderete and Davis 2016*

Methods matter! *more from Alderete Davies 2016*

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

Methods matter! Estimating error frequency

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 different methods, and 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	m̃	ṽ	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

From *Alderete and Davis 2016*

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 by error type

Observations: % of phonotactic violation differs by type, but overall % if 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**

Examples of phonotactic violations

Substitutions (SFUSED record ID # on left)

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.

Deletions

3954 … Lisa, /**Sr**eech and Lisa. (Screech)

8943 … I think you're a /h**u[ŋə**]= hunk-a-rama.

More examples

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

Sequential Blends
4453 … A diary is a /[**sb**]ook xxx a very special book.
5278 … you can't quite /[**pjɪ**rt] xxx put your finger on.
7211 … because we /[spı**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 phonological pattern

Repair strategies: Additions

Onsetless syllables: large percentage of additions (approx. 40%) repair onsetless syllables by filling an onset:

Example: 1170 … that /[baɪ] ^*basically wear all the time.* (I)

Repair strategies: Deletions

Cluster reduction: the majority (73%) of deletions reduce CC clusters, and while they may also produce onsetless syllables, there is a strong tendency for such deletions to be contextual:

Example: 1398 *Stress _e**rr**o**r**s _a**r**e **r**eally* /[_ɛ**r**]. (rare)

Perhaps a role for markedness constraints (OT constraints: Onset, *Complex), even dominated constraints not active in the language (cf. Goldrick & Rapp 2007 on markedness asymmetries)

Results by phonological pattern, cont'd

Sonority sequencing in CC clusters

Onsets

21 of 176 (12%) substitutions and additions into CC onsets violated a sequencing constraint. Ex. *means* → *mreans* (5599)

Codas

4 of 44 (9%) substitutions/additions into CC codas violated sequencing.

Higher incidence of irregularity in CC clusters (sensible, since most phonotactic restrictions apply to clusters)

Tense/lax phonology

Additions in C_CC syllables don't seem to reflect the general avoidance of tense vowels in VCC rimes: of 14 relevant cases, 6 additions had lax vowels, 8 had tense vowels.

Lax vowel: 6	Tense vowel: 8
Eg. *this*[t]	cam[p] [keɪmp]

Effect of collection method on regularity

Summary statistics: 4.34% of errors violate phonotactics, which is far more than 1% reported in prior research. Likely due to robust methods of SFUSED.

SFUSED: collection method is coded for in SFUSED, so we can test for effect of method regularity.

Observation: there is a significant association between collection method and regularity—nearly three times as many phonotactic violations

	Regular	Irregular
Online	660	11 (1.64%)
Offline	1,326	76 (5.55%)

$X2$ = 16.9618, $p < .05$

Discussion: perceptual biases

Dell et al.'s conjecture: 99% standard for regularity too high, will be lower if factor in perceptual biases

SFUSED: offline method and verification protocols have been shown to be more representative of actual occurrence of speech errors, less prone to bias. Lower phonological regularity in SFUSED is direct confirmation of Dell's conjecture.

Room for debate: what's regular/irregular (*θw, gw*)

No room: method affects counts of regularity, same phonotactics used for both offline and online data.

Revised standard: 94-95% regularity

Discussion: new standard and models

Dell et al. (1993): simple recurrent network developed to model errors. Trained on a sample of English words and tested against a set of phonological patterns characteristic of speech error patterns (given below).
→Tried to account for phonological effects, without tactic frames.

- Phonological regularity effect (our focus here)
- Consonant-vowel category effect
- Syllabic constituent effect
- Initialness effect (a.k.a. the word onset asymmetry)

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**, which compares with our findings in natural speech.

Limitations: model trained only on three segment words, designed to account for limited phonological structure, cf. many polysyllabic words in our database. Very likely that Dell et al. also used different phonotactics. Didn't account for prosody (stress errors), non-native sounds (except maybe geminates), and other structures. However, many of the errors with phonotactic violations produced by the SRN resemble the phonotactically illicit errors we have found, with illegal clusters and initials.

Discussion: is there a role for grammar?

Issue: while it appears Dell et al's SRN is a good fit with the data, there still could be a role for grammar in explaining more subtle patterns than categorical regular/irregular combinations.

Goldrick 2011, Goldrick & Daland 2009: speech errors may produce marked structures, but statistically they **tend to reduce markedness**, and markedness is distinct from frequency.

Example: Goldrick & Rapp 2007
Case study of BON, brain-damaged subject with deficit in post-lexical phonological processes (sensitivity to prosody and features)

- More accurate with onsets (96%) than codas (91%)
- More accurate with coronals /t d/ (93%) than dorsals /k g/ (86%)

At least onset/coda difference could not be attributed to frequency effects.

Conjecture: markedness may account for asymmetries that can't be accounted for with pure frequency; if so, supports a notion of gradient constraints now ubiquitous in phonological theory, e.g., Harmonic Grammar, MaxEnt Grammar, Gradient Symbol Processing

Discussion: Evidence for markedness asymmetries?

Warning: *somewhat conjectural, numbers still a bit rough*

Repairs — *from before*
High percentage of additions repair onsetless syllables (40%). Evidence for Onset?
Very high percentage of deletions reduce CC onsets or codas (73%). Evidence for *Complex?

Additions into onsets and codas, by position and context sensitivity

Onsets

	C1	C2	Simple
Contextual	15	61	91
Noncontextual	10	6	10
Total - All	193 (71.2%)		
Totals, N.C.	26		

Codas

	C1	C2	Simple
Contextual	6	16	46
Noncontextual	1	2	6
Total - All	78 (28.8%)		
Totals, N.C.	9		

Observation: many more additions into onset position than codas, which is expected if errors structured by markedness; weaker in noncontextual errors.

Discussion: Evidence for markedness asymmetries?

Warning: *somewhat conjectural, numbers still a bit rough*

Deletions in onsets and codas, by position and context sensitivity

Onsets

	C1	C2	Simple
Contextual	4	28	35
Noncontextual	1	10	5
Total - All	83 (58.9%)		
Totals, N.C.	16		

Codas

	C1	C2	Simple
Contextual	8	6	25
Noncontextual	3	3	11
Total - All	58 (41.1%)		
Totals, N.C.	17		

Observation: still more errors in onsets, but magnitude is far less (cf. 71.2% vs. 28.8% with additions from above); noncontextual errors roughly equal. Expect a difference from additions, because deletions remove segments (good in codas, bad in simple onsets).

Future work: markedness asymmetries

Markedness asymmetries: assemble a range of potential markedness asymmetries and document their patterns in SFUSED

Check for output biases based in frequency: then try to determine if the asymmetry could be explained as an output bias of some kind: segment frequency, sequence frequency, onset/rime frequency, etc.

Anticipating outcomes

- If find markedness asymmetry that can't be explained as an output bias, then we have evidence for a role for gradient constraints.
- Lacking markedness effects independent from frequency effects, then we can account for the speech error patterns with just a SRN like Dell et al's model.
- What I think is likely: mixed bag, some markedness effects strong, others not. Syllabic effects seem strong, but pilot testing of segment markedness (e.g., coronals vs. dorsals) has shown no effect. (Stemberger 1991 explicitly argues for combining frequency effects with other structural patterns, so also has a mixed model.)

Take Homes

New standard for regularity: errors of phonological encoding are phonotactically regular about 94-95% of the time. Phonotactic violations are more common that reported in prior research.

Methodology: new standard is documented using different collection methods; so lower standard is likely the result of perceptual bias.

Grammar in phonological encoding? SRN does a relatively good job of accounting for the data, but markedness asymmetries may emerge that require a role for gradient constraints. *Stay tuned!*

Contributors to SFUSED

Director/analyst/data collector: John Alderete

Research associates
Alexei Kochetov (Toronto)
Stefan A. Frisch (USF)

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

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

Future projects

SFUSED is designed as a multi-purpose database and we actively seek out new project proposals.

New projects can improve the data quality through data cleaning, and also extend the database by introducing new fields.

The database can be imported to Tableau and then explored much faster and deeper.

Projects can combine questions from linguistic theory (e.g., ‘psychological reality of X’) or experimental paradigms.

Contact: alderete@sfu.ca