



John Alderete, Simon Fraser University

Data collectors/analysts: Holly Wilbee, Monica Davies, Olivia Nickel, Queenie Chan, Macarius Chan, Jennifer Williams, Julie Park, Rebecca Cho, Bianca Andreone, Dave Warkentin, Crystal Ng, Gloria Fan

Take Homes

- Supports prior work showing that speech errors respect phonotactic constraints.
- However, quantitative analysis shows that speech errors are less phonologically regular than commonly assumed (<95.7% regularity, cf. 99%).
- Differences in phonological regularity are likely due to perceptual biases in the collection of speech errors.
- Supports the view that respect for phonotactics can be intrinsic to language production processes, and not necessarily the result of phonological grammar.

Context

- Common theme in speech error research: speech errors respect grammatical constraints (e.g., Fromkin 1971, Bock 2011, Stemberger 1983)
- Consensus in classic work that speech errors respect phonotactics (Wells 1951, Fromkin 1971, Boomer and Laver 1968, Nootboom 1967, Garrett 1980), with some even arguing for active phonological constraints
- Quantitative assessment from large speech error based on strong methods found 37 phonotactic violations in 6,300 errors, or 0.59% (Stemberger 1983). Used as a standard that speech errors are phonologically regular 99% of time.

Motivation

- While sound errors do tend to respect phonotactics, other types of speech errors exhibit rampant violations of grammar:
 - Majority of **word shifts** in SFUSED-English are ungrammatical, e.g., *I'm not staying /ə all /up night.* (sfusedE1817)
 - **Sentence blends** frequently ungrammatical: *... in life /I'd never will understand#* (sfusedE348)
 - **Articulatory study** argues phonotactics not a good way to assess lawful speech because most sublexical errors are not of 'alphabetic' symbolic categories (Mowrey MacKay 1990)
- It seems possible that reported 99% standard is too high, because the perceptual systems of data collectors may regularize violations, or they may simply not hear these errors (see Dell et al. 1993).

Methods**SFUSED-English****Simon Fraser University Speech Error Database**

Data collection (cf. Chen 1999 for similar methodology):

- Based in audio recordings of spontaneous speech, typically 2-4 talkers
- Many data collectors (currently 13 data collectors) to reduce researcher bias
- Training regime: one month of training for data collection (phonetic training, speech error identification feedback on listening tests)
- Verification process: all data verified by at least one other person that is distinct from the original data collector; reduces false positives.

Empirical tests of reliability

- Alderete and Davies 2016 document a number of differences between online (traditional observation) and offline collection from audio recordings, consistent with claim offline collection is less prone to bias:
 - Offline is more diffuse, less concentrated in very salient patterns, e.g., blends and exchanges
 - Offline collects more errors from faster speech
 - More noncontextual and uncorrected errors offline
 - Substitutions are more asymmetrical offline
 - More violations of the category constraint offline
- Some errors in SFUSED collected online, can actually test for effect of collection method – see below.

English Phonotactics

Onset Peak Coda

(s)(C1)(C2) X4 (X5) (C6)(C7)(C8)(C9)

- All C positions are optional.
- Banned C1: *ŋʒ*, Banned Codas: *h, j, w*.
- Onset clusters: obstruent + sonorant
- Appendix + C, C always a voiceless stop, *sfrare/loans*
- Banned onset clusters: vd fric/affricate + sonorant, labial + w, coronal nonstrident + l, *θw fV, fVll sr sh gw stw skl*
- On glide *j*: part of peak because of limited distribution, but cannot occur in CC₂ cluster.
- Coda clusters X5 + C6: falling sonority (r > bnasals > obstruents) and *s + p tk; 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.
- Stressed and final syllables are bimoraic (lax vowels occur in closed syllables) and all syllables maximally trimoraic (syllables tense vowels only have simple codas)

Effect of collection method on phonological regularity $\chi^2 = 16.9618, p < .05$

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

Results by Error Type**Substitutions** (SFUSED record ID # on left)*Illicit onsets/appendices*

- 4708 ... pulled back the /zlip/= sleeping bag and ...(sleeping)
- 1500 ... by the maps at the ^selection /fkrin/ (screen)
- 4725 ... about every ^Xbox /srjee ^sixty game that ... (three)
- 5731 ... yeah straight up xxx /tfrjaight up cold. (straight)
- 5739 ... they shoot, /zju shoot The Thick of It ... (you)

Illicit codas/rimes

- 1245 ... Their HOV /laurj xxx lane is like one driver (lane)
- 2223 Let's see if we can heavy /onpa/ shift Susan (NP).
- 5898 Vin Diesel got kicked off of /Rei[n]deer Games ... (reindeer)

Nonnative sounds

- 5035 Which is maybe one of the /beI/ lines ever (best).
- 5964 ... first of all, Katrina /kly/= clearly defined (clearly)

Additions*Illicit onsets, appendix + onset*

- 49 ... get the Ferrari down a /fju xxx few ^floors? (few)
- 1236 ... Haha /θl]isten to you.
- 1248 ... teaching me how to /pfly a ^plane. (fly)
- 1278 I don't like the ^/viral ^marketing. (viral)
- 4187 Yeah, you said ^/pwhy don't you ^put it on one ... (why)
- 5545 ... probably /aft[sjer I ^saw her (after)
- 5599 ... talking a ^dream, what that ^dream /fmr]eans ... (means)
- 6642 ... holding it with his /b]weird ^blue hand. (weird)

Deletions*Illicit codas/rimes*

- 1526 The ^person /kermip ^up to the desk.
- 7682 ... /bjat/ a down side that (but)

Deletions

- 3954 ... Lisa, /Sreech and Lisa. (Screech)
- 8943 ... I think you're a /hulnə/= hunk-a-rama.

Exchanges

- 4581 ... the children in the trailer for /Moon[rarj /Keez= Moonrise Kingdom.

Sequential Blends

- 4453 ... A diary is a /s]book xxx a very special book.
- 5278 ... you can't quite /p]irt xxx put your finger on.
- 7211 ... because we /s]itk xxx we, we speak film

Word Blends

- 870 ... /sastæ/ makes me frisky. (pasta, sauce)
- 7120 Top ten /thways to make me cry (things, ways)
- 7270 ... /sofm-bwan-di xxx uh in the ... (someone, somebody)

Summary Statistics

| Type | N | Violations | % |
|-------------------|--------------|------------|-------------|
| Substitutions | 1,376 | 44 | 3.20 |
| Additions | 358 | 33 | 9.22 |
| Deletions | 169 | 3 | 1.78 |
| Exchanges | 37 | 2 | 5.41 |
| Shifts | 7 | 0 | 0.0 |
| Sequential Blends | 57 | 4 | 7.02 |
| Word Blends | 72 | 4 | 5.56 |
| Total | 2,076 | 90 | 4.34 |

Results by Phonological Pattern**Repair strategies**

- Onsetless syllables: large percentage of additions repair onsetless syllables by filling an onset
 - 1170 ... that /ba/ ^basically wear all the time. (I)
- 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:
 - 1398 \$Stress \$errors \$are \$really \$[_er]. (rare)

Sonority sequencing in CC onsets and codas

- 21 of 176 (12%) substitutions and additions into CC onsets violated a sequencing constraint
- 4 of 44 (9%) substitutions/additions into CC codas violated sequencing
 - Higher incidence of irregularity in CC clusters.

Tense/lax phonology

- Additions in CC codas 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.

Concluding remarks**Are phonotactics intrinsic to phonological encoding?**

- Dell et al. 1993 showed that model of phonological encoding trained on a sample of English words made errors that respected English phonotactics, but not to the high standard set at the time (99% regularity, his: 92-95%)
- More representative sample given here is rather close to Dell et al.'s model predictions; indeed, probably closest to 94% when online errors removed.
- Potential role for grammar: markedness asymmetries (Goldrick 2011, Goldrick and Daland 2009), where observed patterns not predicted by output biases based in raw frequencies.