Supplemental Material to 'How Does Dialect Exposure Affect Learning to Read and Spell? An Artificial Orthography Study'

Phonotactic rules for Levenik, based on restrictions on English phonotactics (Harley, 2006; Crystal, 2003) and our additional rules for avoiding ambiguity

Williams, G. P., Panayotov, N., & Kempe, V.

<–! Note: Tipa package is used for producing IPA characters in LaTeX for .pdf output –>

We included a number of rules in our string generation alogirthm that ensure words are generated from our subset of characters while adhering to English phonotactics. These are as follows:

1. All phonological words must contain at least one syllable, and thus must contain at least one vowel and can have a maximum of three consonants in an onset, and four in a coda; for simplicity.
2. Sequences of repeated consonants are not possible.
3. The velar nasal /ŋ/ never occurs in the onset of a syllable.
4. The glottal fricative /h/ never occurs in the coda of a syllable.
5. The affricates /dʒ/ and /tʃ/ and the glottal fricative /h/ cannot occur in an onset with any other consonant.
6. The first consonant in a two-consonant onset must be an obstruent, such as the plosives, e.g. [p, t, k, b, d, g]; the fricatives, e.g. [f, s, ʃ, x, v, z, ʒ, ɣ]; or affricates, e.g. [tʃ] and [dʒ].
7. The second consonant in a two consonant onset must not be a voiced obstruent, i.e. 7 sounds cannot be produced, e.g. [b, d, g, v, ð, z, ʒ, or dʒ].
8. If the first consonant of a two consonant onset is not an /s/ it must be a liquid or a glide, e.g./l/, /ɹ/, /w/, or /j/. The only time the second consonant can be a voiceless obstruent or nasal is when the first consonant is an /s/. The only voiceless obstruent that cannot occur after /s/ is /ʃ/. For our purposes, this means:

- No [m], /n/, and [l] in the first part of a complex onset (i.e. CC. . .).
- No [b] or [d] in the second part of a complex onset (i.e. CC. . .).
- If the first consonant in a complex onset is not /s/ then the second must be /l/.

9. The substring rule: Every subsequence of consonants contained within a bigger sequence must itself obey all the phonotactic rules. This is only relevant to onsets with 3 consonant clusters, but also applies to codas with 3 or more consonants.
10. No glides in syllable (i.e. 2 or more consonant) codas (i.e. no /w/ or /j/). Aside from this, the first consonant can be anything but /h/ (which does not occur in codas).
11. The second consonant in a two-consonant coda must not be /ŋ/, /ð/, /ɹ/, or /ʒ/ (and are subsequently disallowed in the third and fourth positions due to the substring rule). Co-occurrence relations also dictate which pairs of consonants can go together in a coda. Below, we have summarised those relations for the items in our current study, note that below shows the items that can come first and those that can follow on in the second position of the coda:

- /m/ – /p/, /d/, /z/
- /n/ – /θ/, /s/, /t/, /tʃ/, /dʒ/, /d/, /z/
- /s/ – /t/, /k/, /p/
- /k/ – /s/, /t/

- /b/ – /d/, /z/
- /d/ – /z/
- /f/ – /θ/, /s/, /t/
- /l/ – /θ/, /s/, /t/, /k/, /p/, /tʃ/, /dʒ/, /d/, /z/, /ʒ/, /f/, /v/, /m/, /n/

12. If the second consonant in a complex coda is voiced, so is the first (i.e. if CC is [b], [d], [m], [n], [l] then CC must be [b], [d], [m], [n], or [l].)
13. When a non-alveolar nasal is in a coda together with a non-alveolar obstruent, they must have the same place of articulation, and the obstruent must be a voiceless stop. For us, if we have [m], the next segment must be [b].
14. Two obstruents in the same complex coda must share voicing: this means [s], [k], and [f] can go together, and [b] and [d] can go together.

There are lesser considerations which do not have explicit rules, such as no onsets in English begin with /dl/ (which introduces a schwa) or /tl/. Other restrictions will apply (e.g. stressed free morphemes). However these will be taken into consideration after the initial generation of our set.

Additionally, we have included the following restrictions:

1. No [ks] combination in the onset or coda in order to avoid ambiguity as to whether this represents one grapheme or two (e.g. x vs. ks).
2. No nasals (i.e. [m], [n]) following the dental alveolar approximant (i.e. [l]); no [lm] or [ln] in complex codas in order to avoid coarticulation of [l] with the nasals resulting in loss of the [l] phoneme.
3. No [sf] combination in the onset or coda because it is very rare in English due to being typically used in loan words.
4. We maximally selected onsets with 2 consonant sounds to avoid further restrictions on the design of words following the substring rule. As Crystal (2003) notes, only the following constructions are possible in English, beginning with /s/ and followed by a stop and approximant:

- /s/ + /m/ + /j/
- /s/ + /t/ + /ɹ/
- /s/ + /t/ + /j/
- /s/ + /p/ + /jɹl/
- /s/ + /k/ + /jɹlw/

Following all of these rules, we are left with the following options for two consonant onsets: [sm], [sn], [sk], [sl], [kl], [fl], [bl]

For the two consonant codas, we are left with the following: [fs], [ls], [lf], [lk], [ld], [sk], [ns], [nd], [bd]

Finally, all words had to adhere to the following sllable structures:

- Monosyllabic: CVC, CCVC, CVCC
- Bisyllabic: CV.CCV, CCV.CV, CCV.CVC

We produced 7 items for each syllable structure, resulting in 42 constructed words.