Phonological Studies, Vol. 23, pp. 77-20

# Speech errors and phonological patterns: Insights from psycholinguistic and linguistic theory

John Alderete
Simon Fraser University

**ABSTRACT.** In large collections of speech errors, phonological patterns emerge. Speech errors are shaped by phonotactic constraints, markedness, frequency, and phonological representations of prosodic and segmental structure. While insights from both linguistic theory and psycholinguistic models have been brought to bear on these facts, research on the phonological patterns of speech errors rarely attempts to compare analyses from these different perspectives, much less integrate them in a coherent system. This article investigates the phonological patterns in the SFU Speech Error Database (SFUSED) with these goals in mind. In particular, it examines the impact of language particular phonotactics on speech errors and the role of linguistic representations for syllables and tone. The empirical findings support a model that includes both production processes impacted by frequency and explicit representations of tone from phonological theory.

**Keywords**: speech errors, language production, methodology, phonotactics, tone, English, Cantonese

## 1. Introduction

The scientific study of speech errors is chock-full of facts of interest to phonologists. That is, when examining sufficiently large collections of speech errors, certain phonological generalizations emerge. For example, there is the syllable position constraint, according to which sounds slip into the same syllabic positions as their source words (Boomer and Laver 1968; Fromkin 1971), as *leading list* for *reading list*. Sound errors also exhibit the phonological similarity effect (Cutler 1980; Dell and Reich 1981), in that intended and error sounds tend to be phonologically similar. Another pattern is the repeated phoneme effect (Dell 1984; MacKay 1970): spoonerisms like *heft lemisphere* (from *left hemisphere*) illustrate this tendency for sound errors to share a phonological context (e.g., _ε) in both the intended and source words.

The shape of speech errors also supports fundamental assumptions in phonology. For example, the single phoneme effect states that most sound errors involve a single segment, and not sequences or features (Nooteboom 1969; Shattuck-Hufnagel 1983), giving psychological reality to phonological segments. Perhaps the constraint with the most direct connection to phonology is the phonological regularity effect, according to which speech errors are phonologically regular, i.e., obey language particular phonotactic rules (Stemberger 1983; Wells 1951). Clearly, speech error patterns are rich in phonological patterns.

Facts such as these have led to a number of converging views in contemporary phonology and language production research. That is, while all recognize important differences between the two enterprises, a common ground has emerged concerning certain assumptions about phonological structure. Both phonological theory and language production recognize the existence of phonological segments: they are indispensable to the business of describing phonological distributions and processes in phonology, and they are also a fundamental unit of speech planning (Fromkin 1971). Indeed, some types of speech errors, like sound exchanges, are nearly impossible to account for without recognizing phonological segments (Dell 2014). Similarly, both enterprises make explicit use of syllable structure. Syllables are critical to all manner of segmental and suprasegmental processes (Blevins 1995; Itô 1989), and they have an equally important role in language processing. In models of language production, whole syllables are generally represented in the mental lexicon because they produce a syllable frequency bias, and individual segments are associated with syllable roles (Dell 1986; Levelt et al. 1999). Some models even formalize the

retrieval of entire syllables because of the importance they play in languages, like Mandarin, with reduced syllabaries (Chen 2000; O'Seaghdha et al. 2010). Finally, similarity structure has been crucial to the analysis of harmony and disharmony phenomena in phonology (Frisch 1996; Frisch et al. 2004), and it is no less important to language production because a host of speech production processes are sensitive to phonological similarity (Fricke et al. 2016; Goldrick 2004).

This common ground has supported in some cases to an even greater rapprochement of the two disciplines. That is, some models have gone beyond the uses of phonological structure discussed above and directly employed constructs from generative phonology in the analysis of language production processes. The use of syllable frames to account for phonological regularity is a good example. Phonological encoding is the process of selecting and assembling the phonological segments of an already activated syntactic word. It is often assumed that this selection process involves inserting sounds into slots in a syllable frame (e.g., Dell 1986; Shattuck-Hugnagel 1979). The syllable frame, like other structures in generative linguistics, represents the productive capacity of a talker in terms of possible and impossible syllables. It provides a structure for allowable onsets, nuclei, and codas, and the insertion of sounds must respect these syllabic roles. Speech errors obey phonotactics, therefore, because both correct and incorrect encoding of sounds (the latter resulting in errors), is constrained by syllable frames. Indeed, syllable frames account for phonotactic effects in ways entirely parallel to their account of phonotactics in synchronic grammars (Itô 1986; Levin 1985).

Branching out to other problems, it is not hard to find additional ways in which insights from phonological theory have informed production models. In phonology, feature specification has played a role in explaining when certain sounds are active in phonological derivations (Archangeli 1984; Kiparsky 1982; Paradis and Prunet 1991). Stemberger (1991) argues that a similar role for feature specification can explain certain asymmetries in the production of marked (specified) relative to their default (unspecified) counterparts. In another parallel, selected segments need to be ordered serially in phonological encoding, and this is accomplished in part by an alignment of these activated segments with a metrical frame for the word. The alignment of these two structures in the WEAVER++ model (Levelt and Wheeldon 1994) uses a left-to-right mapping procedure that is essentially identical to the linking of segment and prosodic structure in Autosegmental Phonology (Goldsmith 1990). A more contemporary contribution comes from the role of markedness constraints in Optimality Theory (McCarthy and Prince 1995; Prince and Smolensky 1993/2004). Markedness constraints lie at the heart of phonological analyses in Optimality Theory because they provide the principal motivation for unfaithful mappings, or phonological processes. Recent work in language production has argued that these markedness constraints shape the outcome of speech errors stochastically by favoring unmarked structure over marked structure (Goldrick 2011; Goldrick and Daland 2009).

It is important to emphasize that these applications of phonological theory are not simply convenient assumptions made in order to implement a model. Rather, they involve building on a core assumption in phonological theory, and illustrating that the same insight has a critical role in the analysis of language production processes.

While the exchange of theoretical insights across disciplines is in principle a good thing, it can in some cases lead to competing explanations of the same phenomenon. Language production models have very different underlying assumptions than generative models of grammar, and offer solutions to the problems discussed above that draw on these inherent assumptions rather than phonological theory. Thus, contemporary models of production (Dell 1986; Levelt et al. 1999; Smolensky et al. 2014; Vousden et al. 2000; Warker and Dell 2006) model speech production

within an activation dynamics in which linguistic units like sounds or features have an activation value that determines in part how it is encoded. This activation dynamics predicts behavior with numerical computations of large bodies of data, and it is highly interactive and sensitive to the frequency of sounds and words. Generative models of phonology, such as Lexical Phonology (Kiparsky 1985) and Optimality Theory (Prince and Smolensky 1993/2004), in general are not dynamical, interactive, or sensitive to frequency effects, and they tend to employ symbolic computation rather than numerical computation. In other words, production models just work very differently than phonological grammars, and so it is not a surprise that they produce very different analyses of speech error patterns.

To flesh out some of these differences, many production models have a frequency bias that stems from core assumptions. For example, in the spreading-activation model of Dell (1986), and much subsequent work built on this model, phonological encoding has a bias for high frequency sounds because these sounds are connected to many more words in the mental lexicon, and they receive activation from them when they occur in the same context. This frequency bias makes similar predictions to markedness because unmarked sounds tend to also be high frequency sounds, which leads to a competing explanation. The same reasoning can address why sound errors that involve more than one segment tend to be valid sub-syllabic units, as in attested complex onsets or syllable rimes. These units are also high frequency bigrams or trigrams, and so, because they are linked up to actual words, frequent sequences likewise have an output bias. This extended role of frequency again supports a competing explanation of phonotactic regularity, because models can be trained to capture the frequencies of sequences, but they lack an explicit representation of syllables to enforce phonotactics (Dell et al. 1993).

This article reviews a set of phonological generalizations in speech errors and attempts to reconcile these converging views and competing explanations. In particular, given the differences between language production models and generative models, we ask if all direct roles for constructs from phonological theory are valid and appropriate. An important theme of this article concerns the empirical basis for phonological generalizations in speech errors. In the next section, we review the methods and results of a new large database, the Simon Fraser University Speech Error Database (SFUSED), with these generalizations in mind. It turns out that new data from SFUSED English reveals a weaker role for phonotactics in speech errors, and this in turn supports accounts of phonological regularity given by language production models (section 3). We also examine new data from SFUSED Cantonese to investigate the question of how tone is represented and encoded in phonological encoding (section 4). By contrast, the nature of tone slips in Cantonese supports a converging view whereby explicit representations of tone are accessed in phonological encoding, contrary to the dominant view in language production.

## 2. The Simon Fraser University Speech Error Database (SFUSED)

The database includes two data collections, SFUSED English, with 10,104 speech errors in English (Alderete 2018), and SFUSED Cantonese, with 2,549 Cantonese errors (Alderete and Chan 2018). The SFUSED was developed as a multi-purpose database designed to support research in both linguistics and psycholinguistics. Speech errors are documenting in rich detail in the database. Each error is analyzed for a set of 80 word properties, sound properties, and major and special class features that cross-classify the error within a standard taxonomy of errors (Stemberger 1993). The development of SFUSED Cantonese is part of a larger effort to give rich linguistic detail to speech errors in non-Indo-European languages, which have been understudied in language production research (Costa et al. 2007; Griffin and Crew 2012).

Speech errors are unintended, non-habitual deviations from a speech plan (Dell 1986). Sound errors, for example, are pre-conscious substitutions, deletions, or additions of sounds that stray from the normal pronunciation of the intended word. They are not dialectal or idiolectal features, which are habitual behaviors, or changes of the speech plan mid-utterance. In creating the SFUSED, we developed a set of methods for collecting errors from audio recordings using this definition to avoid a range of problems in data quality and reliability. The full details of these methods are documented in Alderete and Davies (2019), but we summarize the key assumptions that relate to the data discussed below.

Our methods involved three important assumptions that mark a departure from many prior speech error studies. First, we collected errors primarily from audio recordings of natural conversations. In particular, we collected audio recordings from eight podcast series freely available on the Internet. The podcast series we listened to had to meet certain criteria, including high production quality, large amounts of natural and unscripted speech, a gender balance among the talkers, and an exclusion of trained media professionals. The data was therefore collected "offline" in the sense that the conversation had taken place in the past and in another location. This approach contrasts from many prior studies that used an "online" collection method (Fromkin 1971; Stemberger 1982/1985), i.e., collecting errors by making on-the-spot observations in spontaneous speech.

Second, our approach used multiple data collectors that had undergone about a month of training. The current team includes 16 undergraduate and graduate students, plus the author. The students were trained in phonetic transcription, casual speech phenomena, speech error characteristics, and techniques for accurately documenting errors. The latter training involved about three weeks of listening exercises in which the student was asked to collect errors from pre-screened recordings, and then given feedback on their work. This approach contrasts with prior research, which has tended to use large numbers of untrained data collectors, or just one or two highly skilled psycholinguists. Finally, we separate data collection from the classification of the speech error. That is, data collectors submitted errors using a consistent coding scheme, and then these errors were vetted by a data analyst. The analyst re-listened to the audio recording of each error, checked each error against a set of reference material, and then, if deemed a true error, it is incorporated into the database by filling in the field values for the error.

Our attention to these methodological practices is motivated by prior research that has documented a host of problems in data quality and reliability in speech error research (Bock 1996; Ferber 1995). Speech errors are often missed by human listeners, and about a quarter of all errors collected by trained experts are found not to be errors (Cutler 1982; Ferber 1991). Collection by human listeners is also subject to constraints on human perception, and a number of researchers have identified perceptual biases that have the effect of skewing the distribution of errors in large collections (Frisch and Wright 2002; Pérez et al. 2007; Pouplier and Goldstein 2005). For example, collections are skewed toward errors that are more phonetically salient and errors that occur in salient positions, like word-initial contexts.

Our methods were designed to mitigate the impact of these problems. The use of audio recordings reduces perceptual biases for the simple reason that collectors can replay and slow down speech if it is not clear on the first pass, and the acoustic attributes can be investigated with speech analysis software. Audio recordings also allow the analyst to verify idiolectal features and casual speech phenomena, which is critical to verification.

The use of multiple data collectors mitigates perceptual bias because if a specific data collection has a bias, its impact is limited to that individual. Also, each speech error is associated

with the person who collected the error, so if any biases exist, they can be identified and corrected for. Using many data collectors for the same conversation is also necessary because any single collector can only detect about one third of the errors in a speech stream (Alderete and Davies 2019; Ferber 1991). When more than one listener collects data from the same source, they collectively produce a sample that has better coverage of the actual population of speech errors. Finally, by separating data collection and classification, this enables the verification of the error against the reference material on dialectal and idiolectal features and casual speech. Typically, about 25% of the errors are tossed out because they do not meet the definition of a speech error, which underscores how important this extra step is.

In Alderete and Davies (2019), we show in some detail how these methodological considerations affect data composition. In the early stages of our work, we experimented with online data collection, using the same techniques as prior studies to collect roughly 1,100 speech errors. We compared the patterns found in those online errors with a balanced set of speech errors collected offline from audio recordings, and then documented a host of statistically significant differences between the two datasets. For example, it is known that mis-pronunciations of place of articulation are easier to detect than errors in voicing (Cole et al. 1978; Stemberger 1992). We found that errors collected online reflected this bias, having approximately three times more place errors than voicing errors, but that errors collected offline did not reflect the bias. Likewise, errors collected online seem to be far more skewed by the attention bias than offline errors. The rate of exchange errors, like *heft lemisphere* (*left hemisphere*), is a commonly used measure of this bias because exchange errors are highly salient and typically impede on communication (Stemberger 1982/1985). Alderete and Davies (2019) found that the rate of exchanges in offline errors was about 0.38%, and was significantly lower than the 5.6% rate found in online errors. Some error collections using many untrained data collectors have rates of exchange errors exceeding a third or even half of the entire corpus (Dell and Reich 1981; Pérez et al. 2007), which gives a sense of the scale of the impact of this perceptual bias. In summary, the methods for collecting speech errors have a significant impact on the frequency distributions of error patterns. We ask now if this finding has an impact on the characterization of phonotactic patterns.

## 3. Competing explanation: Phonological regularity in SFUSED English

If one is interested in a role for phonological insights in language production, phonological regularity and phonotactic effects is probably the best place to start. Speech errors have long been known to respect phonotactic constraints (Boomer and Laver 1968; Garrett 1980; Nooteboom 1969; Wells 1951). Furthermore, the phonological regularity enforced by phonotactics lies at the heart of phonological analysis. Phonological grammar, through its use of syllable templates, constraints on feature composition, and metrical parsing, formalizes phonotactic regularity. If speech errors reflect phonotactics, it would seem that ideas from phonological theory would have some utility in shaping speech errors.

However, there is also reason to proceed with caution. While there is general agreement in the field that speech errors are phonologically regular, phonotactic constraints are not hard constraints on the output of language production models. That is, it has been known for some time that speech errors do admit phonotactic violations (Hockett 1967; Stemberger 1982). For example, Stemberger's English study showed that roughly 1% of sound errors violated phonotactic constraints, and that production models must admit phonologically irregular forms as well. In addition, it has been conjectured that the lack of reported phonotactic violations in speech error corpora could be due to a perceptual bias against them (Cutler 1982; Shattuck-Hufnagel 1983). Listeners may have regularized them or simply failed to register them as errors. If this is true, then

the actual rate of errors that have phonotactic violations might be much higher, a point we return to below.

In Alderete and Tupper (2018), we examined the role of phonotactics in SFUSED English with these issues in mind. In particular, we examined 2,228 sound errors of various types, including phonological substitutions, additions, deletions, exchanges, and shifts. We also examined sequential blends (e.g, *three fifty* → *thrifty*) and word blends (*trust/faith* → *traith*), because they provide a similar opportunity to test phonotactic rules. We checked the outcomes of each of these error types against a set of established phonotactic rules of English. These rules were based on a standard set of syllable-based phonotactic rules from the literature (Giegerich 1992; Jensen 1993). They essentially state a set of conditions on valid onsets, nuclei, and rimes (see the appendix of Alderete and Tupper (2018) for the explicit system). This phonotactic system gives a categorical assessment: the error was either regular (no phonotactic violations) or irregular (violated phonotactics). While some research has shown that phonotactic regularity is gradient in nature (Frisch et al. 2000), our use of a categorical assessment was motivated by a need to compare our results with prior research that also used a categorical assessment (Dell et al. 1993; Stemberger 1982). Furthermore, as will become clear, our findings are not affected by this assessment procedure.

The results revealed new facts of interest concerning the phonological structure of speech errors. First, the occurrence of phonotactic violations in speech errors is much higher than the 1% rate documented in Stemberger's work. Of the 2,228 sound errors we examined, 102, or 4.58% of them were phonologically irregular, and a search conducted on a larger dataset had a 5.5% rate of phonotactic violations. Thus, there are approximately five times more phonotactic violations in our corpus than the 1% occurrence rate that has been the standard for prior research.

This higher rate of phonotactic violations raises the question, asked originally in Cutler (1982), if the lower rate in prior research is due to perceptual biases against phonotactically illicit forms. This question was addressed in Alderete and Davies (2019) by comparing the rate of phonotactic violations in a balanced set of errors collected offline and those collected online. Recall that the basic finding from this work was that online "on-the-spot" data collection, which was used in Stemberger's study, is more prone to perceptual bias than offline data collection. Alderete and Davies (2019) found that errors collected offline had three times as many phonotactic violations than those collected online, and that rates of violations were significantly associated with collection method. Importantly, this result does not depend on the way phonotactic violations were assessed. The offline and online errors were subjected to the same system of phonotactics, and a statistically significant difference was found. It seems clear, therefore, that this is another instance where the methods matter in collecting speech errors.

In prior research, speech errors were thought to be regular 99% percent of the time (Stemberger 1982). In such a context, the role of phonotactics seems so obviously necessary that no one bothered to consider if the rate of phonotactic violations (1%) could have been explained by chance. Our new findings, however, raise the question of whether phonologically regular forms actually occur above chance levels. For example, in the production of a word like *blue*, there is a chance rate that an error with this word will be phonologically regular, as in *plue*, and a chance rate that it will be irregular, for example, *vlue*. Given that phonological irregular forms have been found to be much higher, we may ask if the rate of these violations in the corpus actually deviates significantly from their chance levels. Using a standard permutation test from Dell and Reich (1981), we examined this question in complex onsets and found something rather unexpected, shown below in Table 1. Substitutions and additions into the second position of a complex onset,

as in *dream* → *dweam*, did in fact occur at above chance levels. However, errors into the first position of an onset, as in *blue* → *plue*, did not. Indeed, additions into the first slot in the error, for example, *last* → *flast*, were phonologically regular 62% of the time in our corpus, which was actually below the 64% chance rate.

**Table 1** *Actual and randomly generated regularity (Table 4 from Alderete and Tupper 2018)*

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

In other words, there appears to be a role for phonotactics in the second position of a complex onset, but this does not seem to be the case in the initial slot. Theories of phonological encoding thus require an analysis of this context-specific effect of phonotactic regularity.

What are the model implications of these findings? Phonotactic regularity has not been extensively studied as a result of model assumptions, that is, using specific mechanisms to predict precise rates of phonologically regular and irregular forms. Because of the near categorical respect for phonotactics documented in Stemberger (1982), many researchers may not have looked beyond the syllable frame account sketched in the introduction. However, one study, Dell et al. (1993), examined this question in some detail with computational modelling. This study is of particular interest here because it investigated the necessity of syllable frame constraints by constructing a model of phonological encoding that did not have them, and testing it against known data. In particular, they developed a connectionist network called simple recurrent network that, through learning, associates a plan representation with a sequence of phonological segments, that is, the pronunciation of a phonological word. The model is sequential in the sense that it outputs a single segment at a time, in the order they occur in the target word. The model is also recurrent in the sense that the current segment is processed in tandem with a knowledge of prior segments. Like other simple recurrent networks (Elman 1990), the network stores the output of the last pass through the network in a state layer, and uses that state layer representation to predict the next segment. In this way, the model in essence learns the valid sequences of English, and it is able to use this connectionist phonotactic system to account for phonological regularity.

To review the results, Dell et al. (1993) trained the model on a sample of English words and varied a host of model parameters, including the frequency of the English vocabulary and the nature of the state layer that encoded the memory of past segments. Under rather reasonable assumptions, they found that the errors produced by this network were phonologically regular about 96.3% of the time, but that the range of regular errors under other parameters was approximately between 89 and 96%. Viewed from Stemberger's 99% standard for regularity, this result seems somewhat off the mark. However, the modeling results are a rather good fit with the findings from SFUSED English, in which the speech errors were regular between 94.5% and 95.5% of the time. It should be noted that the rate of phonologically regular errors in the SFUSED may be slightly lower, given that the methods, while robust to perceptual bias, are nonetheless affected by them. However, Dell's modeling setup only examined simple monosyllabic words, and one might expect higher rates of phonotactic violations in polysyllabic words and words with more complex phonological structure. In summary, the modeling results of the simple recurrent network seem to be an excellent fit of the SFUSED English data, even without an explicit syllable frame for enforcing phonotactics.

In addition to this overall fit of the data, the simple recurrent network also has a rather natural account of the asymmetry between the impact of phonotactics in initial versus non-initial positions. Recall that substitutions and additions have an above chance rate of phonological regularity in a non-initial position, but this is not the case in an initial position. The simple recurrent network was also tested on this initialness effect to give a general account of the so-called word onset bias in phonological encoding. Word onsets are in general more prone to error (Shattuck-Hufnagel 1987; Wilshire 1998), with many studies having between 50% and 90% of all sound errors in this position. Dell's simple recurrent network was shown to exhibit a comparable initialness effect as a direct consequence of the sequential nature of the network. In non-initial positions, the network has a memory of what has come before and can use this information to predict future segments. But this is not the case word initially because there are no prior segments. This basic feature of the model is also consistent with our finding that phonotactic constraints in complex onsets are only in force in the second consonant of the cluster. Given that the vast majority of these are also word-initial, as in *flast*, the simple recurrent network predicts this behavior as well. Future work will have to investigate initialness in other contexts, but the sequential nature of the model also seems to make exactly the right predictions for SFUSED English.

Let's return to the larger question of the role of phonological theory in language production processes. On the one hand, we have an account of phonological regularity based in syllable frames outlined in the introduction. This model in essence accounts for phonological regularity in the same way that phonological grammars do, with an explicit template describing the productive capacity for phonological words. There are two problems with this approach. First, it requires an explicit account for phonotactic violations, because, nothing else said, it predicts 100% regularity. This does not seem like an insurmountable problem because many contemporary theories of grammar include random noise in computing outcomes (Goldrick and Daland 2009; Jäger 2004), and so it seems likely that syllable frames can be made to produce more variable outcomes. However, a more pressing problem is the independent support for syllable frames. They seem to be only necessary to account for phonological regularity in speech errors, and even that requires refinement, given that phonological regularity is not a hard constraint.

The alternative approach using a simple recurrent network addresses these problems. Phonotactic regularity in this model arises naturally from the need to associate a plan with its phonological representation. The training regime needed to learn these associations can thus be thought of as a part of word learning, a developmental process that is necessary in any analysis of language production. In addition to this, the model makes very detailed predictions about the rate of phonological regularity and the specific positions where we expect more errors, and these predictions turn out to be the correct ones. It seems clear, therefore, that the competing explanation offered by this theory is superior to the one employing syllable templates.

**4. Converging view: Encoding tone in SFUSED Cantonese**

Another potential common ground between phonological analysis and language production is the characterization of the planning units in phonological encoding. Phonological encoding creates a speech plan by retrieving and selecting the phonological contents of a word. It turns out that many of the phonological structures that are indispensable to phonological analysis also have an important role in phonological encoding. Phonological segments and other sub-syllabic units are established planning units in many production models (Fromkin 1971; Stemberger 1983), and other phonological structures like syllables and phonological features, while not actively selected in some models, have additional roles in phonological encoding (Chen 2000; Dell 1986). The

characterization of these planning units is therefore another natural context to examine the role of phonological structure in language production models.

An important distinction to make is between units that are directly retrieved and those that are indirectly retrieved (Dell 1990; Garrett 1975; Shattuck-Hufnagel 1979). In phonological encoding, phonological segments are generally assumed to be indirectly accessed and mediated by an activation dynamics that associates selected segments with a syllable frame or a plan representation (Dell 1986; Stemberger 1990). Metrical structure, on the other hand, is retrieved directly, essentially binding activated syntactic words (lemmas) to a metrical representation (Levelt et al. 1999; Meyer and Belke 2007). This distinction is motivated by form preparation experiments that show how prosody facilitates word-form retrieval (Roelofs and Meyer 1998), and it also accounts for a fundamental distinction between speech errors involving segments and errors involving stress. Sub-lexical errors are generally mis-selections of segments, but speech errors involving stress are exceedingly rare and have alternative explanations (Cutler 1980). This contrast is accounted for if segments are actively selected in phonological encoding, because errors arise as mis-selections. Metrical stress, on the other hand, is not actively encoded, and thus cannot be mis-selected (see, e.g., Chen et al. 2002).

Given these assumptions, how is tone accessed in phonological encoding? After all, tone has similarities with both phonological segments and metrical prosody. Tone is like segments in that it has a lexical role in many languages, bringing about contrast in otherwise identical words (Yip 2002). Yet tone is also suprasegmental, generally represented in phonological analysis on a plane distinct from consonants and vowels (Goldsmith 1990). This question is actually at the center of an active debate in psycholinguistics. The dominant view in language production research is that tone in languages like Mandarin Chinese is retrieved directly, much like metrical structure. The evidence for this position is based largely on the findings of Chen (1999), who argued that tone slips in Mandarin are extremely uncommon and the rare cases that exist can be re-analyzed in ways that do not involve mis-selections of tone. Contemporary models of phonological encoding in Chinese languages have thus mapped tone directly to a prosodic frame representation, parallel to the mappings of metrical structure (Chen et al. 2002; Roelofs 2015). A viable alternative exists, however, that treats tone on par with segments. In this view, tone is actively selected in phonological encoding and can be mis-selected in exactly the same ways segments can be. Evidence for this position comes from studies of tone languages in which tone slips are relatively common and can be perseverated, anticipated, and exchanged, just like segments (Gandour 1977; Shen 1993; Wan and Jaeger 1998). In sum, there is currently no consensus about the empirical status of tone slips, nor the model assumptions designed to account for them.

This question was taken up in a recent study of tone errors in Cantonese, a language with six lexical tones (Alderete et al. 2019). The data for this study came from SFUSED Cantonese (Alderete and Chan 2018). This data collection employed the same methods described above, enlisting four native speaker data collectors to collect speech errors from natural conversations. Data collectors were also trained in the correct characterization of Cantonese tones, including screening potential errors for tone mergers, a change in progress that constitutes a habitual behavior and thus cannot be a speech error (Bauer et al. 2003; Mok et al. 2013). All speech errors were vetted by a data analyst that checked them against our reference materials for idiolectal and dialectal variations, again excluding habitual behavior from the set of true errors.

The basic finding was that tone errors are not uncommon in Cantonese, and that they exhibit a host of intricate behaviors that patterns with segmental errors. We documented 432 simple tone errors, i.e., substitutions of tone that supplant the lexical tone, and an additional 236 complex errors

that involved both segmental and tone mis-selections. The simple tone errors alone were the second most common type of speech error, and constituted 20.55% of all sound errors involving a single phonological category. Furthermore, the majority of tone slips (roughly 76%) were found to be contextual in the sense that the intruder tone was identical to a nearby tone, as in the anticipatory error /dou25 jan21 ge33/ → dou33 jan21 ge33 '… (verbal particle) people (sentence particle) … 到人嘅'. This is important because if tone is selected in phonological encoding, we expect anticipatory or perseveratory activation of a nearby tone to lead to such errors, just as they do in segmental errors.

Our 2019 study also investigated a subtle form of evidence that tone is retrieved indirectly. One of the hallmarks of indirectly accessed structure is the existence of interactive spreading effects, or the observation that a structure is not encoded in isolation but part of a larger web of interconnected structure (Dell 1986). For example, as explained in the introduction, the repeated phoneme effect shows that segments are more likely to be mis-selected when they share a phonological environment with a source sound. This effect arises from the activation that flows between the shared phonological structure and the word level, which demonstrates the lack of independence of these structures (Dell 1984; Dell 1988).

We reasoned that, if tone is in fact part of phonological encoding, then we would expect to find a parallel "repeated tone effect". We found such evidence in three distinct data sets. In particular, we found that phonological substitutions are over-represented when the syllables containing the intended and source sounds share a tone, a direct tonal parallel to the repeated phoneme effect. Second, we found that lexical substitutions also interacted with tone because substitutions in monosyllabic words had a greater than chance probability of sharing a tone, supporting a similar observation made for Mandarin Chinese in Wan and Jaeger (1998). Finally, we found tone-to-tone substitutions also exhibit non-independence because the more similar two tones are, the more likely they are to substitute in a speech error. Similarity effects like this in segments are generally analyzed as feedback from phonological features linked to segments (Dell 1986), and so this too supports another kind of non-independence of tone in phonological encoding. In sum, tone slips in Cantonese exhibit a set of interactive spreading effects that has direct parallels in speech errors involving segments (Fay and Cutler 1977; Nooteboom 1969; Shattuck-Hufnagel 1979), further supporting the comparison between tone and segmental encoding.

We have again examined a question, the encoding of tone, that compares a phonological account that recognizes explicit tonal representations with a more processing account that directly maps tone to a prosodic frame. At issue is whether the tonal categories of Cantonese are recognized as planning units or not. After reviewing the arguments from Alderete et al. (2019), it is clear that tone patterns like segments in speech errors and therefore must be recognized as viable planning units in languages like Cantonese. The larger argument is therefore in support of a converging view, identifying another linguistic structure that is indispensable to both phonological analysis and phonological encoding in online production processes.

## 5. Concluding remarks

This article contributes to an on-going discussion about the utility of phonological ideas in models of language production (Fromkin 1971; Goldrick 2002; Goldrick 2011; Stemberger 1983). On the one hand, there appear to be many theoretical insights that inform production models: phonological segments, features, and constituents all seem to have non-trivial roles in the analysis of production processes. On the other hand, it does not seem to be the case that all theoretical innovation in phonology has a natural place in production models. Adoption of these ideas,

therefore, requires detailed examination of alternative explanations that derive more naturally from processing assumptions.

One of the take homes of this article is that methods are extremely important to the correct characterization of phonological patterns in speech errors, and that model implications need to be evaluated from solid empirical ground. The investigation of both phonological regularity in SFUSED English and tone slips in SFUSED Cantonese revealed new facts of critical importance to language production models. Furthermore, the differences between the empirical generalizations established in the SFUSED and prior research stem largely from differences in methods for collecting and analyzing speech errors.

One of these differences has to do with the use of syllable frames as a mechanism for modeling the productive capacity of speakers. Empirical investigation of the impact of phonotactic rules showed that speech errors respect these rules much less than acknowledged in prior research. Also, the impact of phonotactics depends in part on the position of the error within a word, as initial slots are impacted less by phonotactics. These new empirical findings led us to question the role of syllable frames in phonological encoding, and argue for a processing explanation (Dell et al. 1993), because it gives a very natural account of these generalizations and yet it lacks syllable frames.

Our study of tone slips in Cantonese also supported new conclusions within a contemporary debate about how tone is encoded. We found in SFUSED Cantonese that, contra prior research, tone errors are not at all uncommon and exhibit all of the hallmarks of indirectly retrieved phonological structure. That is, they are largely contextual and are selected within a production system that interacts with other sound and word structures. These findings support a model in which tone is actively selected, just like consonants and vowels, and one that is a radical departure from mainstream models in which tone is encoded like stress (Chen 1999; Roelofs 2015).

While these conclusions are supported empirically, they do not reduce the importance of related phonological constructs in other aspects of production. Thus, rejection of syllable frames does not mean that syllable structure is irrelevant to phonological encoding, or language production processes in general. Rather, syllable structure is pervasive in both domains. For example, syllables are argued to be indirectly accessed in Chinese languages in some models (O'Seaghdha et al. 2010). Syllabic roles must also be aligned with segments to account for the general fact that slips respect these roles (Warker and Dell 2006), and syllables themselves must be aligned with tone structure (Alderete et al. 2019). Syllables program units have also been proposed as a link between word-form retrieval and articulatory processing (Levelt et al. 1999). In sum, syllable structure still supports a converging view in phonology and psycholinguistics in many aspects of production processes.

Finally, we must also qualify the above results by emphasizing that phonological and language production accounts are not always in competition with each other. In particular, there seem to be some phenomena that require ideas from both phonology and production models. Thus, accounts of consonant substitutions are sometimes portrayed as a competition between markedness and frequency accounts, because unmarked phonological structure tends to be high frequency, and speech errors in large collections tend to produce outcomes that reduce markedness and increase frequency. However, reviews of these effects in many languages emphasize the necessity of both markedness and frequency (Buchwald 2005; Goldrick 2011; Romani et al. 2017). In other words, the role of frequency in production models does not really supplant the role of markedness, or vice versa, because both of these factors appear to be necessary in some contexts.

## References


Alderete, John. 2018. *Simon Fraser University Speech Error Database - English (SFUSED English Beta)*. Burnaby, British Columbia, Canada.

Alderete, John & Queenie Chan. 2018. *Simon Fraser University Speech Error Database - Cantonese (SFUSED Cantonese 1.0)*. Burnaby, British Columbia, Canada.

Alderete, John, Queenie Chan & H. H Yeung. 2019. Tone slips in Cantonese: Evidence for early phonological encoding. *Cognition* 191.103952.

Alderete, John & Monica Davies. 2019. Investigating perceptual biases, data reliability, and data discovery in a methodology for collecting speech errors from audio recordings. *Language and Speech* 62.281–317.

Alderete, John & Paul Tupper. 2018. Phonological regularity, perceptual biases, and the role of grammar in speech error analysis. *WIREs Cognitive Science* 9.e1466.

Archangeli, Diana. 1984. *Underspecification in Yawelmani phonology and morphology*: Massachusetts Institute of Technology Doctoral dissertation.

Bauer, Robert S., Kwan-hin Cheung & Pak-man Cheung. 2003. Variation and merger of the rising tones in Hong Kong Cantonese. *Language Variation and Change* 15.211–25.

Blevins, Juliette. 1995. Syllable in phonological theory. The handbook of phonological theory, ed. by J. Goldsmith, 206–44. Cambridge, MA: Blackwell.

Bock, Kathryn. 1996. Language production: Methods and methodologies. *Psychonomic Bulletin and Review* 3.395–421.

Boomer, Donald S. & John D. M. Laver. 1968. Slips of the tongue. *International Journal of Language and Communication Disorders* 3.2–12.

Buchwald, Adam. 2005. *Representation, repair, and well-formedness: Grammar in spoken language production*: Johns Hopkins University Doctoral dissertation.

Chen, Jenn-Yeu. 1999. The representation and processing of tone in Mandarin Chinese: Evidence from slips of the tongue. *Applied Psycholinguistics* 20.289–301.

—. 2000. Syllable errors from naturalistic slips of the tongue in Mandarin Chinese. *Psychologia* 43.15–26.

Chen, Jenn-Yeu, Train-Min Chen & Gary S. Dell. 2002. Word-form encoding in Mandarin Chinese as assessed by the implicit priming task. *Journal of Memory and Language* 46.751–81.

Cole, Ronald A., Jola Jakimik & William E. Cooper. 1978. Perceptibility of phonetic features in fluent speech. *Journal of the Acoustical Society of America* 64.45–56.

Costa, Albert, R.-Xavier Alario & Núria Sebastián-Gallés. 2007. Cross-linguistic research on language production. The Oxford handbook of psycholinguistics, ed. by G. Gaskell, 531–46. Oxford: Oxford University Press.

Cutler, Anne. 1980. Errors of stress and intonation. Errors in linguistic performance: Slips of tongue, ear, pen, and hand, ed. by V. Fromkin, 67–80. New York: Academic Press.

—. 1982. The reliability of speech error data. Slips of the tongue and language production, ed. by A. Cutler, 7–28. Berlin: Mouton.

Dell, Gary S. 1984. Representation of serial order in speech: Evidence from the repeated phoneme effect in speech errors. *Journal of Experimental Psychology: Learning, Memory and Cognition* 10.222–33.

—. 1986. A spreading-activation theory of retrieval in sentence production. *Psychological Review* 93.283–321.

—. 1988. The retrieval of phonological forms in production: Tests of predictions from a connectionist model. *Journal of Memory and Language* 27.124–42.

—. 1990. Effects of frequency and vocabulary type on phonological speech errors. *Language and Cognitive Processes* 5.313–49.

—. 2014. Phonemes and production. *Language, Cognition and Neuroscience* 29.30–32.

Dell, Gary S., Cornell Juliano & Anita Govindjee. 1993. Structure and content in language production: A theory of frame constraints in phonological speech errors. *Cognitive Science* 17.149–95.

Dell, Gary S. & Peter A. Reich. 1981. Stages in sentence production: An analysis of speech error data. *Journal of Verbal Learning and Verbal Behavior* 20.611–29.

Elman, Jeffrey. 1990. Finding structure in time. *Cognitive Science* 14.179–211.

Fay, David & Anne Cutler. 1977. Malapropisms and the structure of the mental lexicon. *Linguistic Inquiry* 8.505–20.

Ferber, Rosa. 1991. Slip of the tongue or slip of the ear? On the perception and transcription of naturalistic slips of tongue. *Journal of Psycholinguistic Research* 20.105–22.

—. 1995. Reliability and validity of slip-of-the-tongue corpora: A methodological note. *Linguistics* 33.1169–90.

Fricke, Melinda, Melissa Baese-Berk & Matthew Goldrick. 2016. Dimensions of similarity in the mental lexicon. *Language, Cognition and Neuroscience* 31.639–45.

Frisch, Stefan A. 1996. *Similarity and frequency in phonology*: Northwestern University Doctoral dissertation.

Frisch, Stefan A., Nathan R. Large & David S. Pisoni. 2000. Perception of wordlikeness: Effects of segment probability and length on the processing of nonwords. *Journal of Memory and Language* 42.481–96.
Frisch, Stefan A., Janet Pierrehumbert & Michael B. Broe. 2004. Similarity avoidance and the OCP. *Natural Language and Linguistic Theory* 22.179–228.
Frisch, Stefan A. & Richard Wright. 2002. The phonetics of phonological speech errors: An acoustic analysis of slips of the tongue. *Journal of Phonetics* 30.139–62.
Fromkin, Victoria. 1971. The non-anomalous nature of anomalous utterances. *Language* 47.27–52.
Gandour, Jack. 1977. Counterfeit tones in the speech of southern Thai bidialectals. *Lingua* 41.125–43.
Garrett, Merrill. 1975. The analysis of sentence production. The psychology of learning and motivation, Advances in research and theory, Volume 9, ed. by G.H. Bower, 131–77. New York: Academic Press.
—. 1980. Levels of processing in sentence production. Language production: Vol. I. Speech and talk, ed. by B. Butterworth, 177–220. London: Academic Press.
Giegerich, Heinz J. 1992. *English phonology: An introduction*. Cambridge: Cambridge University Press.
Goldrick, Matthew. 2002. *Patterns of sound, patterns in mind: phonological regularities in speech production*: Johns Hopkins University Doctoral dissertation.
—. 2004. Phonological features and phonotactic constraints in speech production. *Journal of Memory and Language* 51.586–603.
—. 2011. Linking speech errors and generative phonological theory. *Language and Linguistics Compass* 5.397–412.
Goldrick, Matthew & Robert Daland. 2009. Linking speech errors and phonological grammars: Insights from Harmonic Grammar networks. *Phonology* 26.147–85.
Goldsmith, John. 1990. *Autosegmental and metrical phonology*. Oxford: Blackwell.
Griffin, Zenzi M. & Christopher M. Crew. 2012. Research in language production. The Cambridge handbook of psycholinguistics, ed. by M.J. Spivey, K. McRae & M.F. Joanisse, 409–25. Cambridge: Cambridge University Press.
Hockett, Charles F. 1967. Where the tongue slips, there slip I. To Honor Roman Jakobson, Vol II. The Hague: Mouton.
Itô, Junko. 1986. *Syllable theory in prosodic phonology*: University of Massachusetts, Amherst Doctoral dissertation.
—. 1989. A prosodic theory of epenthesis. *Natural Language and Linguistic Theory* 7.217–59.
Jäger, Gerhard. 2004. Maximum entropy models and stochastic Optimality Theory. University of Potsdam.
Jensen, John T. 1993. *English Phonology*. Amsterdam: John Benjamins.
Kiparsky, Paul. 1982. From cyclic phonology to lexical phonology. The structure of phonological representations, ed. by H.v.d. Hulst & N. Smith, 131–75. Dordrecht: Foris.
—. 1985. Some consequences of Lexical Phonology. *Phonology* 2.85–138.
Levelt, Willem J. M., Ardi Roelofs & Antje S Meyer. 1999. A theory of lexical access in speech production. *Behavioral and Brain Sciences* 22.1–75.
Levelt, Willem J. M. & Linda Wheeldon. 1994. Do speakers have access to a mental syllabary. *Cognition* 50.239–69.
Levin, Juliette. 1985. *A metrical theory of syllabicity*: Massachusetts Institute of Technology Doctoral dissertation.
MacKay, Donald G. 1970. Spoonerisms: The structure of errors in the serial order of speech. *Neuropsychologia* 8.323–50.
McCarthy, John J. & Alan Prince. 1995. Faithfulness and reduplicative identity. University of Massachusetts Occasional Papers 18, Papers in Optimality Theory, ed. by J. Beckman, S. Urbanczyk & L. Walsh, 249–384. Amherst, MA: Graduate Linguistic Student Association.
Meyer, Antje S. & Eva Belke. 2007. Word form retrieval in language production. The Oxford handbook of psycholinguistics, ed. by G. Gaskell, 471–87. New York: Oxford University Press.
Mok, Peggy P.-K, Dongui Zuo & Peggy W.-Y Wong. 2013. Production and perception of a sound change in progress: Tone merging in Hong Kong Cantonese. *Language Variation and Change* 25.314–70.
Nooteboom, S. G. 1969. The tongue slips into patterns. Leyden studies in linguistics and phonetics, ed. by A.J. van Essen & A.A. van Raad, 114–32. The Hague: Mouton.
O'Seaghdha, Padraig G., Jenn-Yeu Chen & Train-Min Chen. 2010. Proximate units in word production: Phonological encoding begins with syllables in Mandarin Chinese but with segments in English. *Cognition* 115.282–302.
Paradis, Carole & Jean-François Prunet (eds) 1991. *The special status of coronals: Internal and external evidence*. San Diego: Academic Press.
Pérez, Elvira, Julio Santiago, Alfonso Palma & Padraig G. O'Seaghdha. 2007. Perceptual bias in speech error data collection: Insights from Spanish speech errors. *Journal of Psycholinguistic Research* 36.207–35.
Pouplier, Marianne & Louis Goldstein. 2005. Asymmetries in the perception of speech production errors. *Journal of Phonetics* 33.47–75.

Prince, Alan & Paul Smolensky. 1993/2004. *Optimality theory: Constraint interaction in generative grammar*. Malden, MA: Blackwell.
Roelofs, Ardi. 2015. Modeling of phonological encoding in spoken word production: From Germanic languages to Mandarin Chinese and Japanese. *Japanese Psychological Research* 57.22–37.
Roelofs, Ardi & Antje Meyer, S. 1998. Metrical structure and planning in production of spoken words. *Journal of Experimental Psychology: Learning, Memory and Cognition* 24.922–39.
Romani, Cristina, Claudia Galuzzi, Ceclia Guariglia & Jeremy Goslin. 2017. Comparing phoneme frequency, age of acquisition, and loss in aphasia: Implications for phonological universals. *Cognitive Neuropsychology* 34.449–71.
Shattuck-Hufnagel, Stefanie. 1979. Speech errors as evidence for a serial-ordering mechanism in sentence production. Sentence processing: Psycholinguistic studies presented to Merrill Garrett, ed. by W.E. Copper & E.C.T. Walker, 295–342. Hillsdale, NJ: Erlbaum.
—. 1983. Sublexical units and suprasegmental structure in speech production planning. The production of speech, ed. by P.F. MacNeilage, 109–36. New York: Springer Verlag.
—. 1987. The role of word onset consonants in speech production planning: New evidence from speech error patterns. Motor and sensory processes of language, ed. by E. Keller & M. Gopnik, 17–51. Hillsdale, NJ: Erlbaum.
Shen, Jiaxuan. 1993. Slips of the tongue and the syllable structure of Mandarin Chinese. Essays on the Chinese language by contemporary Chinese scholars, ed. by S.-C. Yau, 139–61. Paris: Centre de Recherche Linguistiques sur L'Asie Orientale. École des Hautes Études en Sciences Sociales.
Smolensky, Paul, Matthew Goldrick & Donald Mathis. 2014. Optimization and quantization in gradient symbol systems: A framework for integrating the continuous and the discrete in cognition. *Cognitive Science* 38.1107–38.
Stemberger, Joseph P. 1982. The nature of segments in the lexicon: Evidence from speech errors. *Lingua* 56.235–59.
—. 1982/1985. *The lexicon in a model of language production*. New York: Garland.
—. 1983. *Speech errors and theoretical phonology: A review*. Bloomington: Indiana University Linguistics Club.
—. 1990. Word shape errors in language production. *Cognition* 35.123–57.
—. 1991. Radical underspecification in language production. *Phonology* 8.73–112.
—. 1992. The reliability and replicability of naturalistic speech error data. Experimental slips and human error: Exploring the architecture of volition, ed. by B.J. Baars, 195–215. New York: Plenum Press.
—. 1993. Spontaneous and evoked slips of the tongue. Linguistic disorders and pathologies. An international handbook, ed. by G. Blanken, J. Dittmann, H. Grimm, J.C. Marshall & C.-W. Wallesch, 53–65. Berlin: Walter de Gruyter.
Vousden, Janet I, Gordon D. A. Brown & Trevor A. Harley. 2000. Serial control of phonology in speech production: A hierarchical model. *Cognitive Psychology* 41.101–75.
Wan, I-Ping & Jeri J Jaeger. 1998. Speech errors and the representation of tone in Mandarin Chinese. *Phonology* 15.417–61.
Warker, Jill A. & Gary S. Dell. 2006. Speech errors reflect newly learned phonotactic constraints. *Journal of Experimental Psychology: Learning, Memory and Cognition* 32.387–98.
Wells, R. 1951. Predicting slips of the tongue. *Yale Scientific Magazine* 3.9–30.
Wilshire, Carolyn E. 1998. Serial order in phonological encoding: An exploration of the "word onset effect" using laboratory-induced errors. *Cognition* 68.143–66.
Yip, Moira. 2002. *Tone*. Cambridge: Cambridge University Press.