

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

John Alderete, Paul Tupper

Simon Fraser University

Speech errors involving manipulations of sounds tend to be phonologically regular; they obey the phonotactic rules of well-formed words. We review the empirical evidence for phonological regularity in prior research, including both categorical assessments of words and regularity at the granular level involving specific segments and contexts. Since the reporting of regularity is affected by human perceptual biases, we also document this regularity in a new dataset of 2,228 sub-lexical errors that was collected using methods that are demonstrably less prone to bias. These facts validate the claim that sound errors are overwhelmingly regular, but the new evidence suggests speech errors admit more phonologically ill-formed words than previously thought. Detailed facts of the phonological structure of errors, including this revised standard, are then related to model assumptions in contemporary theories of phonological encoding.

1. Introduction

Speech errors tend to respect grammar. At the sublexical level, for example, sound errors result from mis-selections of phonological planning units, as in ... *blood clot in a certain area of the blain* (*brain*, sfusedE-1419), but the resulting words tend to conform to the phonotactic rules governing phonological well-formedness (Boomer & Laver, 1968; Garrett, 1975; Wells, 1951). Likewise at the word level, lexical substitutions tend to be syntactically well-formed, e.g., *This side has green Vs, I mean red Vs* (sfusedE-797),¹ because intended and error words usually have the same part of speech (Bock, 1982; Garrett, 1975). Slips of tongue also obey rules of allomorph selection, as illustrated by the accommodation of the definite article in *an eating marathon* → *a meeting arathon* (Fromkin, 1971). Finally, at the sentence level, errors like sentence blends and role mis-assignments tend to respect syntactic constraints, as emphasized by the following quotation from (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”.

In many models of language production, grammatical regularity is the consequence of a kind of built-in structural matrix, and therefore not part of what is explained by the model. An example is the spreading-activation theory of Dell (1986). This model combines an activation dynamics in the mental lexicon with a system of tactic frames for slotting selected items into well-formed sentences, words, and syllables. Nodes in the mental lexicon both encode the substance of the linguistic unit, for example, the phoneme [t], and a category label that determines how it is slotted in the tactic frame, as in “Onset” for onset positions. Phonological encoding in this system selects from a restricted subset of phoneme nodes, searching only for phonemes with Onset labels when filling the onset position, and only those with Coda labels

¹ Here and throughout we index examples from SFUSED with their record ID numbers so that they may be explored in more detail in this database.

when filling the coda, etc. The phonological regularity is therefore in a sense pre-compiled in the system. For example, the fact that speech errors never have the velar nasal [ŋ] in onset position is simply a consequence of the assumption that there is no node for [ŋ]/Onset in the lexical network. The model does generate some phonologically unacceptable syllables, but the larger result is that the majority of illegal syllables are systematically excluded by the assumed tactic frames for syllables and category matching. In these contexts, phonological regularity is in a sense a hard constraint on phonological encoding because the model literally cannot produce words like *ŋot*.

The problem with this approach is that grammatical regularity in general is not a hard constraint. While phonotactic violations are rare in sound errors, they do occur with some frequency in natural speech. Stemberger (1983) applied a consistent set of phonotactic rules to a large data set of naturally occurring speech errors in English and found that about 1% of them violate English phonotactics. Branching out to word errors, the error and intended words are usually the same part of speech (e.g., 99% of the time in malapropisms examined in Fay and Cutler (1977)), but category constraint violations do occur and they result in ungrammatical sentences, as in *My small is kinda, my house is kinda small* (sfusedE-4524). Finally, some syntactic errors like word shift are strongly ungrammatical, as in *I'm not staying all up night* (correct order: *staying up all night*, sfusedE-1817). It seems therefore that, at various levels of analysis, language production processes have a strong tendency towards grammatical regularity, but nonetheless do admit irregular forms or ungrammatical sentences in specific contexts.

Regarding phonotactics, Stemberger (1983: 32-33) notes in passing that it is not obvious that the overwhelming fact of phonological regularity entails that phonotactic constraints are directly applied as rules in language production processes. Independently necessary mechanisms that are sensitive to frequency structures could potentially explain the facts of phonological regularity and irregularity. This question was examined in some detail in Dell, Juliano, and Govindjee (1993). These researchers built a computational model of phonological encoding that learns permissible sound sequences of English through error-corrective learning. In particular, they developed a simple recurrent model (Elman, 1990; Jordan, 1986) that learns to associate a plan representation with a sequence of segments in three segment words. As a simple recurrent network, the model processes segments with a memory of prior segments, allowing it to create associations between segment classes in the string and effectively learn permissible sequences. When tested, the model did generate some phonotactically illicit words, but with optimal model parameters the output was phonologically regular approximately 96.3% of time. The important point is their model differed from the spreading activation model in that it lacked a built-in structural matrix for phonotactics. It lacked the tactic frames that guarantee regularity, and yet the learned associations account for the overwhelming fact of phonological regularity.

The success of Dell et al.'s model raises an important issue in psycholinguistics, and speech errors in particular, namely how precisely is grammatical regularity encoded in language production processes. Are grammatical constraints familiar from linguistic analysis directly encoded, as in the tactic frames, or are the effects of constraints derived from more basic principles of the model? This question has been central to many models of language production (Dell et al., 1993; Goldrick, 2002; Stemberger, 1982/1985), and yet there does not appear to be a clear consensus in the field as to if grammatical regularities are directly encoded in production processes, and if so, which ones. Our aim here is to assemble the evidence on regularity in a particular domain, phonological regularity, and relate this evidence to a range of models that have been designed to account for it. We do this by both reviewing the empirical generalizations

that characterize regularity in sub-lexical speech errors, and also contributing new evidence on regularity in English speech errors. In particular, we will investigate phonological regularity in the Simon Fraser University Speech Error Database (SFUSED), a database of naturalistic speech errors collected in a way that is demonstrably less prone to methodological problems found in prior studies. It turns out that the impact of these problems has an effect on the characterization of phonological regularity, and so our revised characterization of regularity informs the models designed to account for it.

In the next section, we review past research on phonological regularity in speech errors, examining both regularity as a categorical classification of words and the individual factors that contribute to this holistic assessment. Section 3 reviews a host of models of phonological encoding, including the connectionist models discussed above, with a focus on how model assumptions predict phonological regularity. Section 4 investigates a set of sub-lexical errors in SFUSED and documents the fact that phonological irregularity, or speech errors with phonotactic violations, is more widespread than assumed in past accounts. The final discussion section relates these findings to the models reviewed in section 3 and also sketches some additional questions raised by these findings that can be pursued in future research.

2. Background

2.1 Categorical phonotactics

A common claim in language production research is that speech errors obey phonotactic constraints. Phonotactics is the language particular system of constraints that govern possible contexts for sounds and possible sound combinations. For example, English phonotactics prohibit the velar nasal [ŋ] syllable initially (though languages like Vietnamese and Indonesian freely permit such syllables), and it also bans word-initial [p-s] sequences (contrast with French words like *psychologie*). Phonotactic constraints can be applied gradiently (Kessler & Treiman, 1997) and even learned in experiments (Dell, Reed, Adams, & Meyer, 2000), but claims about the structure of speech errors tend to focus on the categorical distinction between phonologically regular and irregular errors (Dell et al., 1993; Fromkin, 1971; Stemberger, 1983), i.e., whether the errors obey phonotactics (regular) or not (irregular).

One context in which phonotactics have been applied is the analysis of word blends, like *smort*, a blend of *small* and *short* (sfusedE-2061). Blends have the potential to create phonotactic violations because illegal combinations may arise from merging the segments of two words: *smort* could have ended up as *shmort*, with an illegal initial [ʃ-m] sequence. But blends tend not to produce phonotactic violations (Hockett, 1967; Wells, 1951). Likewise, exchanges of two sounds, like *torn korkilla* (for *corn tortilla*, sfusedE-1495), have the potential to violate phonotactics, and yet, when examined, they too are phonologically regular (Fromkin, 1971; Garrett, 1975; MacKay, 1970). Fromkin (1971: 40-41) further argues that phonotactics may account for certain phonological repairs observed in exchanges, as in *play the victor* → *flay the pictor*, in which the illegal onset [v-l] is avoided by devoicing [v] to [f]. Phonological regularity has also been observed in the larger class of sound errors, including sound substitutions, deletions, and additions (Boomer & Laver, 1968; Fromkin, 1971; Nooteboom, 1969). The online processes involved in production planning of segments, or phonological encoding, therefore generally result in phonologically regular and permissible words.

While the above research did not document phonological regularity quantitatively, the implicit assumption in at least some studies is that phonological regularity is largely an absolute requirement. Others, however, acknowledge that phonotactic violations are possible in speech

errors, but just exceedingly rare (Garrett, 1975; Hockett, 1967; Wells, 1951). Hockett (1967: 98), for example, confidently states, “I have no doubt that blending can yield pronunciations with constituent sounds that sound outside the ‘normal’ phonological system of a language”. The first study to rigorously investigate this issue in a large dataset was Stemberger (1983), based on a larger dataset from (Stemberger, 1982/1985). Stemberger applied a consistent set of phonotactic generalizations to approximately 6,300 speech errors and found 37 examples that violated English phonotactics, which is less than 1% of his data (but roughly 1% of just sound errors and blends, which totalled approximately 3,660 in this dataset). All of the violations involved either irregular rimes or illicit consonant clusters, as in *dlorm* for *dorm*. These numbers, together with the careful methods employed in this study, support the contention that speech errors are overwhelming regular, but they do in fact exhibit phonotactic violations.

A methodological question that arises in this context is if the rarity of phonologically irregular speech errors is due to the way they are collected. Speech error collection is notoriously difficult and subject to well-known perceptual biases (Bock, 1996; Pérez, Santiago, Palma, & O’Seaghdha, 2007). Some have conjectured that the relative rarity of phonotactic violations could be due to perceptual bias in data collectors against hearing words that violate phonotactics (Cutler 1982; Dell et al. 1993; Shattuck-Hufnagel 1983; cf. Stemberger 1983). Section 5 examines phonological regularity in the SFUSED database, and this methodological issue is examined in some detail.

2.2 Phonological markedness and frequency

The regular versus irregular distinction represents a global assessment of the overall well-formedness of an utterance. Speech errors are either categorically regular or irregular, and this opposition does not distinguish among various shades for irregularity. However, both contemporary phonology and psycholinguistics analyze regularity at the granular level by recognizing the specific structures and contexts of phonotactic generalizations. Thus, a considerable body of speech error research “drills down” to the specific segmental and prosodic structures referred by the phonotactic constraints, giving both a better description of the speech error patterns and a more insightful analysis of the nature of phonological regularity.

An important theoretical concept for investigating regularity in specific structures and contexts is markedness. Markedness evaluates different types of linguistic structure and establishes an ordering among them that distinguishes marked and unmarked elements. A structure is unmarked on a dimension if it represents the default or basic state for the structure in that dimension, and marked structure adds complexity to this default state. Traditionally, marked structure is learned later in language development, eliminated in historical sound change, implies the corresponding unmarked structure in typology, and impaired in language pathologies like aphasia (Greenberg, 1966; Jakobson, 1941; Trubetzkoy, 1939). For example, the ejective velar [kʰ] are marked relative to their unmarked counterpart [k] because it adds the physiological complexity of a glottalic airstream mechanism to a velar articulation. It also exhibits all of the hallmarks of marked structure: it is learned later, eliminated in change, and implied in linguistic typologies (i.e., the existence of [kʰ] implies [k] in sound inventories). One line of research has investigated the impact of markedness constraints in the structure of speech errors. For example, Blumenstein’s classic (1973) study of aphasics investigated consonant substitutions in which the intended and intruder consonants differ only in a single feature (e.g., [b] and [p] for [±voice]). In natural conversations of 17 aphasics, this study found a greater than chance tendency for marked consonants to be replaced by their unmarked counterparts (in Broca’s and Wernicke’s aphasics, but not conduction aphasics).

An equally important tool for digging deeper into phonological regularity is frequency, including both type and token frequency. Speech error analysis must reckon with the token frequencies of sounds because sounds with greater token frequency present more opportunities for an error (or have higher “availability”), and so counts of individual patterns need to be scaled to these baselines (Shattuck-Hufnagel & Klatt, 1979). Independent of this, speech errors also may exhibit an output bias for elements with high type frequency: high frequency sounds are less prone to error than low frequency sounds, and low frequency sounds tend to be replaced by high frequency sounds (Dell, 1986; Goldrick & Larson, 2008; Levitt & Healy, 1985; Motley & Baars, 1975). In a tongue twister experiment, for example, Levitt and Healy (1985) showed that high frequency [s] replaced lower frequency [ʃ] significantly more often than the opposite substitution pattern.

These two factors, markedness and frequency, lead to competing explanations of the same data, because unmarked sounds are very often more frequent than their marked counterparts. Thus, [s] is both unmarked relative to [ʃ] and it has greater frequency. The competing explanations offered by markedness and frequency are one of the great unresolved issues in speech error research. On one side of the debate, some studies have proposed clever designs that distinguish the effects of markedness and frequency and shown markedness effects distinct from frequency (Goldrick, 2002; Goldrick & Rapp, 2007; Romani, Olson, Semenza, & Granà, 2002). Others have taken the opposite position, either arguing that frequency shapes error patterns and markedness does not (Motley & Baars, 1975; Stemberger, 1991a), or that markedness does not shape error patterns and instead error patterns emerge largely from baseline frequency, or error opportunity (Shattuck-Hufnagel & Klatt, 1979). Still others have taken the position that both markedness and frequency predict speech error patterns and language production needs a mechanism for combining their effects (Buchwald, 2005; Kupin, 1982).

The effects of markedness and frequency have been examined in a variety of different phonological structures. Single consonant substitutions are by far the most common type of speech error, so it is not a surprise that many studies of naturalistic speech errors have explored these effects with consonant confusions (Blumenstein, 1973; Broeke & Goldstein, 1980; Shattuck-Hufnagel & Klatt, 1979; Stemberger, 1991a). Speech errors involving vowels tend to be less common, and studies investigating vowel substitutions, e.g., Shattuck-Hufnagel (1986), do not investigate these questions. Tone has been investigated in Mandarin by Wan and Jaeger (1998), who found that the specific substitution patterns do not support a role for type frequency or markedness, but instead reflect baseline frequencies. Traditionally, markedness has played an important role in the analysis of syllable structure. A number of studies have examined the structural changes in errors and related them to the same phonotactic constraints developed in contemporary phonology for syllable structure related processes, including consonant cluster resolution and onset formation (Béland & Paradis, 1997; Blumenstein, 1973; Buchwald, 2009), epenthesis (Béland, 1990; Buchwald, 2005; Buchwald, Rapp, & Stone, 2007), constraints on coda consonants (Béland, Paradis, & Bois, 1993; Goldrick & Rapp, 2007), and sonority sequencing (Romani & Calabrese, 1998). While unresolved issues pertaining to the viability of markedness and frequency remain, their effects and interaction have provided the basis for constructing many concrete hypotheses for predicting the phonological structure of speech errors.

2.3 Frequency and anti-frequency

The effects of frequency have also been weighed against a kind of regularity related to the specification of linguistic structure, often called an “anti-frequency” bias because it predicts

patterns that run counter to those predicted by frequency. In naturally occurring speech errors, Shattuck-Hufnagel and Klatt (1979) documented a set of asymmetrical specific consonant substitutions that are unexpected if frequent segments replace infrequent ones: e.g., [s] → [ʃ] is more common than [ʃ] → [s] (cf. Levitt & Healy 1985) and [t] → [tʃ] more often than [tʃ] → [t]. These are unexpected because [s t] are some of the most frequent sounds in English, and the palatals [ʃ tʃ] are some of the least frequent, supporting their claim of the existence of a palatal bias that runs counter to frequency expectations. Stemberger (1991a) examined these and other consonant pairs using the SLIP priming paradigm and confirmed these anti-frequency patterns for palatals and a host of other feature classes in English.

In what sense is anti-frequency a phonological regularity? The basic proposal in Stemberger (1991a) and related work (Stemberger, 1991b, 1992, 2004) is that infrequent sounds are replaced in specific contexts because they are default segments that are specified differently from the intruder sounds. Following a trend in phonological theory (Archangeli, 1984; Kiparsky, 1982; Paradis & Prunet, 1991), every natural class of sounds has a default segment that can be under-specified for predictable or non-contrastive features. Because default segments lack feature structure, and competing sounds in a neighboring word have this structure, the activation dynamics that selects the appropriate feature of a segment tends to favor segments that are fully specified, like [ʃ], which is specified for [-anterior], as opposed to [s], which is not. Stemberger (1991) further shows how this principle of ‘items that lack structure get over-ridden’ accounts for substitutions patterns in five other feature classes and a preference for addition (where the target word lacks a full segment) over deletion. While recent work has questioned the viability of anti-frequency effects (Frisch, 1996; Frisch & Wright, 2002; Pouplier & Goldstein, 2005), the extent of these effects in English phonological encoding, and their relationship with phonological theory and processing more generally (e.g., Lahiri & Marslen-Wilson 1992), suggests that it is a form of regularity that must be contended with.

2.4 Learning phonotactics

Speech errors are an important form of evidence for the existence of phonotactic constraints. The fact that sound errors and blends tend to reflect phonotactics is a subtle and strong measure of a language user’s instincts for possible sound combinations. Because phonotactics are language particular, one might assume that these instincts develop over an entire lifetime, accumulating in adulthood and reflecting all of our language learning experiences. A relatively new line of research has shown that the same subtle measure, respect for phonotactics in errors, can be used to assess phonotactic learning in relatively short time intervals, for example, the amount of time it takes to participate in a psycholinguistics experiment.

Dell et al. (2000) developed the implicit learning paradigm, a variant of the tongue twister technique for inducing speech errors (Wilshire, 1999) in which the experimental stimuli themselves reflect phonotactic patterns that can be contrasted with other phonotactic constraints. It is designed to test the Implicit Learning Hypothesis, which posits that talkers learn phonotactics directly from experience, even in short intervals, and do not require an intention to learn phonotactics. This paradigm contrasts language-wide constraints, like [ŋ] is not possible syllable-initially, with experiment-wide constraints, as reflected in the fact that [s] is in an onset in all of the stimuli. The general finding is that experiment-wide constraints are indeed reflected in the error patterns, and can be learned very quickly, perhaps as quickly as 9-10 experimental trials (Taylor & Houghton, 2005), but they are not as robust as language-wide constraints, which are nearly categorical. Subsequent research showed that the strength of the learning depends in part on phonological similarity (Goldrick, 2004), and that more subtle “second order” constraints

that formalize sound combinations (e.g., [k] only occurs in syllables that have the vowel [ɪ]) can also be learned, but they require more time (J. Warker & Dell, 2006; J. A. Warker, 2013). The import of this research for our discussion of phonological regularity is that it is simply not the case that phonotactics require a lifetime of learning. Phonotactic constraints very much like the ones we find in the world's languages can be learned directly from experience in a relatively short time period.

3. Phonological regularity and models of phonological encoding

Is phonological regularity a product of the design of models of phonological encoding, and if so, what are the specific mechanisms responsible for it? Below we review a variety of language production models and point out key assumptions that relate to phonological regularity, and also models that lack such structure by design.

3.1 Spreading activation models, starting with Dell (1986)

The spreading-activation model of Dell (1986, 1988) accounts for different kinds of regularity with a combination of linguistic and processing assumptions. Like most models discussed here, this model assumed that words and sentences are broken down into the linguistic units employed in contemporary linguistics, e.g., morphemes, syllables, segments, and features for sub-lexical units (see e.g., Fromkin (1971)). In addition to these assumptions, the model posits a set of tactic frames that characterize linguistic productivity at different linguistic levels. For example, syllables have the general template: Onset + Nucleus + Coda. How this template is filled in is determined by spreading activation within a richly interconnected lexical network. Thus, abstract activation is sent from intended words to all linked morphemes, then from morphemes to linked syllables, phonemes, and phonological features. A set of interface rules selects the phoneme nodes with the highest activation level that also conforms to the syllabic role labels associated with the slots in the tactic frames.

Though not developed explicitly to account for phonological regularity, the spreading-activation model accounts for two facts of interest here. First, it characterizes general phonotactic regularity through the interfacing rules. The model does generate some unacceptable syllables (e.g., final open syllables with lax vowels, Dell 1986: 296), but these patterns are not really the result of the design of the model. As a rule, however, only licit onsets are allowed in onset position, licit codas in coda position, etc., because the interfacing rules match the syllable role categories in the tactic frames to the nodes in the lexical network. Second, the model predicts frequency effects as output biases in phonological encoding. In particular, there is a frequent sound bias because frequent sounds are linked to more words in the lexical network, and the processing of non-target units contributes preferentially to frequent sounds, leading to higher overall activation for these sounds over less frequent sounds. In sum, this model predicts categorical respect for most phonotactic constraints and also an output bias for frequent sounds.

The model of phonological encoding developed in Dell et al. (1993) is also designed to account for phonological regularity, but with rather different model assumptions. This model is a simple recurrent network (Elman, 1990; Jordan, 1986) that learns to associate a speech plan with a sequence of segments. It is implemented as a multilayer connectionist network with a context layer that encodes a memory of prior segments, which, through training, learns phonotactic generalizations. Though the model has an activation dynamics, it differs from the above spreading-activation model in many respects: it is a sequential network that outputs segments one at a time rather than as chunks, it has no syllable or sub-syllabic nodes, it is not interactive in the sense that phonological encoding is encapsulated from other production processes (e.g.,

lemma selection), and segments are distributed representations rather than atomic nodes. These assumptions amount to a fundamental difference with the spreading-activation network because it is a model of phonological regularity, but it lacks constructs like a syllable tactic frame specifically designed to govern this regularity.

Dell et al. (1993) tested a number of model parameters and evaluated model performance against four benchmark patterns, including Stemberger's (1983) 99% standard for phonological regularity. Indeed, this is the only account that we are aware of that investigates model performance quantitatively for phonological regularity. The parameters that best fit the data resulted in phonologically regular forms approximately 96.5% of the time. This undershoots the 99% standard, but the authors raise the empirical issue that Stemberger's 99% standard may in fact be too high because of the existence of perceptual biases that mitigate against the detection of phonotactic violations (see sections 2 and 4). While the model did not explicitly examine this, the sequence learner is sensitive to phoneme frequency, so we assume that it will also have an output bias for frequent segments.

A final model worthy of note is the activation dynamics model developed in J. Warker and Dell (2006). This model is in a sense a combination of the two models discussed above, but designed to account for implicit learning of phonotactics. The model accounts for the work of the tactic frames of Dell (1986), but through learned associations in a multi-layer network. In particular, it takes as input a set of segments together with a whole syllable, e.g., /f, s, æ; FAS/. A hidden layer representation learns through training to associate these inputs with segments that have a syllabic role, effectively binding segments with syllable position in the output layer. Tests of the syllabification model using different training protocols shows that it can both learn language wide generalizations, like the fact that [ŋ] can never occur in an onset, and also generalizations from recent experience, like the kinds of phonotactic facts gleaned from implicit learning experiments.

3.2 WEAVER++

Phonological regularity can also be investigated in WEAVER++ (Levelt, Roelofs, & Meyer, 1999; Roelofs, 1992), a feedforward model of spreading activation that accounts for many of the same language production processes as the Dell models discussed above. The empirical focus on this model is principally latency data in speaking, for example reaction time data in picture naming experiments. But its model assumptions do make predictions about the structure of speech errors, including phonological regularity. In word-form retrieval (see Meyer and Belke (2007) for a more detailed summary), the morphemes of already selected lemmas are activated, then a mapping of morphemes to component segments takes place, followed by the selection of whole syllables called syllable program nodes. A separate parallel mapping activates a metrical frame that is used as a serial order mechanism for linking segments up with a prosodic frame. The binding of segments to syllabic roles in WEAVER++ is rather different than syllabification in Warker and Dell (2006). In particular, a procedure called "binding by checking" in which each node has a procedure associated with it that, when the node is activated, checks whether the node is associated with an appropriate node the next level up. For example, in the phrase *red sock*, binding by checking both verifies whether the segment [r] is associated with the lemma <red> and whether the syllable program node [rɛd] is associated with [r].

Speech errors of the kind we are interested in here, namely sound errors, are the result of failures of this binding procedure. For example, the [r] in *blood brother* can be bound to both syllables [brʌd] and [brʌðr], resulting in an anticipation error *brood brother*. Computational modeling of these binding failures in Levelt et al. (1999) demonstrates that WEAVER++

accounts for the relative frequencies of anticipations, perseverations, and exchanges in naturalistic speech errors and also the speed-accuracy trade-off in speech production whereby higher error rates are observed at faster rates (Dell, 1995). Detailed measures of phonological regularity have not been investigated, but it seems plausible that mishaps in binding could produce phonotactic violations, as in *blood tll* (*blood toll*), where the [l] is perseverated to produce an illicit onset [tl]. Because of model assumptions, however, it does not seem that type frequency or transitional frequency will play a role in predicting these errors. The reason for this is that in WEAVER++, segments are listed for every morpheme and so syllabic role binding is a local matter of associating the component segments with a syllable. In other words, type frequency and other factors like phonological markedness or feature specification will not have a role, and so error patterns are predicted directly from baseline frequencies, or error opportunity.

3.3 Scan-Copier model

The Scan-Copier model of Shattuck-Hufnagel (1979) and Shattuck-Hufnagel and Klatt (1979) makes predictions about regularity and frequency that overlap with WEAVER++ model and differs sharply with the Dell models described in 3.1. This model is a theory of serial order in language production that employs a symbolic-computational architecture, and slips of tongue arise from certain malfunctions in the operations of this architecture. A key assumption of the model is dual representations: word representations include a linearly ordered sequence of planning segments, and in speaking, these segments are copied into a series of target slots equal in number to the number of planning segments (compare target slots with tactic frames and metrical frames of the models discussed above). Phonological encoding involves a scan-copier algorithm that selects segments from the intended word's planning segments and copies them over into the target slots. This copying process is monitored by the so-called check-off monitor, which deletes planning segments as they are copied into slots, and the error monitor, which detects and deletes error-like sequences. Sound exchanges like *corn tortilla* → *torn korkilla*, for example, arise because [t] is mis-selected by the scan-copier algorithm as the first segment in *corn*, but correctly deleted once selected by the check-off monitor, leading to subsequent selection of [k] in the second word.

The assumptions of this model predict that type frequency is not a factor in error rates, and that the main factor in predicting error rates is baseline frequency, or error opportunity. The reason for this is that errors occur because of mis-selections by the scan-copier algorithm, and mis-selections are caused by slot and segment similarity. For example, the two stops [b] and [p] are equally similar in features and generally occur in the same slots. The rate of [b] → [p] relative to [p] → [b] therefore is a function of the opportunities to produce [b] relative to [p], or token frequency. It is less clear if the model predicts phonotactic violations, and if so, at what rate. Presumably the notion of slot similarity and the error monitor could be enriched in ways that make clear predictions about this kind of regularity, but we are not aware of any such enrichments that make detailed predictions like the Dell et al. (1993) simple recurrent network model.

3.5 Models with underspecification

As alluded to above, Stemberger (1991) assumes that phonological encoding is subject to both frequency and anti-frequency effects, the latter resulting from the lack of competition from underspecified segments. Thus, in Stemberger's analysis of the palatal bias, according to which palatals (marked, less frequent) tend to replace corresponding alveolars (unmarked, more frequent), as in /s/ → /ʃ/ > /ʃ/ → /s/, the alveolar is actually unspecified for the phonological feature [anterior]. At the featural level, therefore, when a specified palatal must compete with an

unspecified alveolar, the former tends to win for the simple reason that the lack of a structure cannot be activated. Stemberger (1991) makes a parallel between cases like this and the addition bias, which likewise favors specified segments over absent segments (Stemberger & Treiman, 1986). The interaction between frequency and the propensity for default sounds to be overwritten by similar sounds has not been formalized in any detail to determine how the different patterns may play out. However, it is clear from this study that this model actually predicts patterns that run counter to output biases for frequent segments.

3.6 Models with symbolic constraints

The Dell (1986) spreading activation model encodes phonotactic constraints as regularities and conformity to the syllable tactic frames, a kind of symbolic system similar to syllable structure building algorithms of the day (Itô, 1986; Levin, 1985). With the advent of Optimality Theory (Prince & Smolensky, 1993/2004), syllable structure related phonology, and phonotactic constraints in general, have been formalized as the combined impact of a large number of constraints, dubbed markedness constraints, that operate on surface forms. A number of production models have reconceptualised the role of regularity with explicit use of markedness constraints. An early example is the use of syllable structure markedness in the analysis of addition and deletion errors in aphasic speech given in Béland and Paradis (1997). This analysis is built on the Theory of Constraints and Repair Strategies (Paradis, 1988; Paradis & LaCharité, 1993), that, much like Optimality Theory, motivates phonological processes as the need to satisfy certain syllable shape constraints, like constraints against branching onsets. Béland and Paradis examined roughly 700 sound errors in a study of a French speaking patient with primary progressive aphasia and compared productions of words containing marked structures (e.g., complex onsets, coda consonants, words with vowel hiatus) with words that lacked such structures and instead had “perfect” CV syllables. They found significant differences between the productions of syllables containing marked structures and words with CV syllables, and endeavoured to document strong parallels between these error patterns and patterns of loanword adaptations. This program of explaining speech error patterns with markedness constraints familiar from contemporary phonology has been very productive, encompassing both segmental markedness (Goldrick, 2002; Goldrick & Rapp, 2007) and prosodic markedness (Béland, 1990; Béland et al., 1993; Buchwald, 2005, 2009; Romani & Calabrese, 1998).

The assumption that markedness is a key factor in shaping the phonological structure of speech errors leads to a problem for cases that do not improve on markedness. Such patterns are well-attested in all speech error corpora and can be primed for in experiments that elicit speech errors (Levitt & Healy, 1985; Motley & Baars, 1975). Indeed, such mappings are a formal problem for constraint-based models of grammar, like Optimality Theory and the Theory of Constraints and Repair Strategies, because markedness is the only explicit mechanism that can cause phonological changes. As a result, input-to-output mappings always improve on markedness, a property referred to as harmonic ascent (Prince, 2007). Theories of phonological encoding that draw on markedness must therefore include some mechanism for accounting for errors that do not improve on markedness.

There are at least three enhancements to constraint-based theories that address this problem. Boersma and Pater (2008) model language production as the output of a Harmonic Grammar (Legendre, Miyata, & Smolensky, 1990), where the constraint weights associated with constraints are randomly perturbed with each production. For most weight assignments, the correct result is produced, but if weights are perturbed enough so that a constraint has a negative weight, a form is now rewarded for violating it. This can occasionally lead to harmonic ascent. A

similar approach is suggested by Goldrick and Daland (2009), who implement a harmonic grammar with a Hopfield network. In this model, each constraint in the harmonic grammar is represented as a family of connections, each with its own weight. As with Boersma and Pater, speech errors violating harmonic ascent may occur when a weight is perturbed and becomes negative, but this model still predicts that marked to unmarked mappings should occur more often in speech errors than the reverse mapping.

Another theory taking Harmonic Grammar as a starting point is the Gradient Symbol Processing model of Smolensky, Goldrick, and Mathis (2014). Rather than Hopfield dynamics of Goldrick and Daland (2009), stochastic gradient ascent is used to find the most harmonic output for a given input. When dynamics are run under optimal parameter settings, the system inevitably converges to the correct output for the input. However, when these ideal conditions are not obtainable, the network can converge into suboptimal answers. This approach has the advantage that noise is not postulated on top of a perfectly functioning model in order to generate errors, but rather the noise is essential for the functioning of the model under any conditions. A prediction of Gradient Symbol Processing is that output x should be produced roughly according to the probability

$\exp(-H(x)/T)$, where $H(x)$ is the harmony of form x and T is a parameter that determines how random the output is. This result is similar to what may be obtained using a Maximum Entropy grammar to model speech errors (Hayes & Wilson, 2008). All of these harmonic grammar-based solutions have not yet been shown to account for the many detailed features of speech error statistics. For instance, it is not clear how intruders are accounted for in these models, and whether the fact that intruders are most likely to occur within a few syllables of the target is reproduced.

3.7 Serial control model

A model with a quite different pedigree is OSCAR (Vousden, Brown, & Harley, 2000), a model based on theories of human performance on serial ordering tasks. In this model, production is performed by three layers of nodes. The lowest layer consists of a collection of oscillators of different temporal frequencies. Half of the oscillators have different frequencies and are known as the “non-repeating portion”, and the other half all have the same frequency and are called the “repeating portion”. Both portions feed into the so-called phonological context vector. The non-repeating portion of the oscillators generates over time a complex high-dimensional signal in one half of the phonological context vector. This portion is used to encode the position at which a talker is currently speaking in an utterance (see Wayment (2009) for a related mechanism for coding phonological context). The repeating portion generates a periodic signal in the other half of the phonological context vector. This portion, being periodic, is able to encode where in a particular syllable the speaker is. The evolution of the phonological context vector is the control signal that generates a “plan” for a sequence of syllables. Preparing a sequence of syllables for production is done by associating each subsequent state of the phonological context vector with a different phoneme, represented in the next layer which holds the phoneme feature vector. An utterance that is prepared for production can be thought of as a binding between utterance position (non-repeating portion of phonological context vector), position within syllable (repeating portion of phonological context vector), and phoneme (phoneme feature vector).

To actually produce the sequence of syllables, the oscillators in the bottom layer are set to their initial position for the sequence and then are allowed to evolve without interference. Each subsequence phonological context vector is activated in the next layer, and these in turn activate

the correct sequence of phonemes. To introduce errors, noise is added to the model at the point where the correct phoneme is selected. The binding of phonemes with roles in the model is such that either a similar but wrong phoneme is selected (an item error) or a phoneme that is in the same utterance but is linked to a somewhat different phonological context vector is selected by mistake (a movement error). The model correctly predicts that item errors should select phonemes that are close to the intended phoneme in terms of phonetic similarity. It also predicts that in movement errors, the intruding phoneme should usually come from the same syllabic role (since the repeating part of the phonological context vector is identical) and then have a preference for nearer phonemes in the word (so that the non-repeating part of the phonological context error is similar). The model also includes inhibition of just produced phonemes, and this accounts for some of the asymmetries between anticipations and perseverations. Phonotactics, frequency, and the lexicon, however, have no explicit role in OSCAR, and the authors state that error phenomena based on these factors are beyond the scope of the model. Until these factors are explored in more detail, therefore, it appears that OSCAR makes similar predictions as the Scan-Copier model.

4. Phonological regularity in SFUSED

Phonological regularity and phonotactic violations have only been investigated on a large scale in one study, Stemberger (1983, 1982/1985). Our goal is to attempt to replicate Stemberger's observation that phonotactic violations are attested in speech errors but quite rare. We focus on data from SFUSED (the Simon Fraser University Speech Error Database), collected by the first author, because it is a similarly large database of English speech errors. Also, as we demonstrate below, the methods employed in SFUSED for collecting speech are arguably less prone to perceptual bias than other studies, and so it can address concerns expressed in e.g. Cutler (1982) that the low incidence of phonotactic violations are due to these biases.

4.1 Methods

SFUSED is a multi-purpose database designed to probe psycholinguistic effects on language production processes. Its current composition includes 10,104 English speech errors and 2,549 Cantonese speech errors. Each entry in the database is a single speech error, classified with 82 variables designed to sort the errors into a traditional taxonomy of errors (Stemberger, 1993) and probe known psycholinguistic biases.

The methods used to collect errors can be characterized with four methodological decisions and practices. First, following a methodology of Chen (1999), speech errors were collected primarily from audio recordings of natural conversations. These included conversations produced as podcasts by third party sources and freely available on the Internet. The podcasts were pre-screened for a number of properties, including high production quality, large segments of unscripted speech primarily from commentators that are not experienced media professionals, and a balance of male and female talkers. The majority of the talkers have language backgrounds one might characterize as 'Standard American English', i.e., an offshoot of the Midlands dialect of the United States frequently heard in Western North America. While some errors were collected initially using traditional observational techniques (see below), the majority of our speech error data comes from these "offline" sources because of the benefits of data quality and reliability.

Second, the speech errors were collected by a small set of experts who were graduate or undergraduate linguistics students at Simon Fraser University. This strikes a balance between two common approaches to collecting speech errors, i.e., ones using a small number of expert

listeners (Shattuck-Hufnagel, 1979; Stemberger, 1982/1985) or a large number of relatively untrained data collectors (Dell & Reich, 1981; Pérez et al., 2007). By using many expert listeners, we address the data collector bias and talker bias that arise when only one or two people are collecting errors (Dell & Reich, 1981). The SFUSED database also allows for the data collector and talker biases to be investigated because the specific collector and talker are recorded for each error. By collecting speech errors with many collectors (the current roster is 13), we also accelerate the data collection processes, as detecting and documenting speech errors in natural speech is extremely time consuming.

The data collectors undergo about a month of training before they can contribute speech errors to the database. This training involves four hours of reading and practice in phonetic transcription and subsequent testing of this skill to confirm transcription ability. Data collectors also get an hour-long introduction to the facts of speech errors. This introduction explains the standard definition of a speech error as a “unintended non-habitual deviation from the speech plan” (Dell, 1986). To give direct experience with speech errors, trainees are asked to spend a couple of hours in their daily lives listening for errors and writing them down using traditional online observational techniques. In the last part of the training, data collectors are asked to examine a set of three pre-screened audio recordings and collect all of the errors that they hear. They are given instructions on how to use the speech analysis software Audacity to listen and re-listen to speech segments, and also slow down the speech if necessary. The errors submitted by the data collectors are then checked against the benchmark errors and the trainees are given feedback on missed errors, correctly identified errors, and incorrectly identified errors. By the end of the training, which takes about a month, data collectors are able to detect on average 32.3% of the pre-identified errors, which is a marked improvement over data collectors that are not given such training (Ferber, 1995).

A final important practice is verification. Incorporation of speech errors in the database involves two steps: initial submission from a data collector, and subsequent verification and classification by a senior data analyst. The analyst was either the first author or an experienced data collector that was later trained as a data analyst. Verification involves both relistening to the sample to confirm the speech facts, and a careful examination of the error for habitual or intentional behavior that would exclude it as an error. In particular, the errors were checked against a set of 29 casual speech rules and known dialectal and idiolectal features (given our libraries of the idiolectal features of our talkers). If the submitted error could be more explained as a known phonetic process or idiolectal feature, or other known confounds, e.g., errors of ignorance, mid-utterance change of the speech plan, slips of ear (Bond, 1999; Vitevitch, 2002), then the error was excluded from the database and not analyzed. Approximately 25% of the submitted errors were excluded in this way, which is consistent with the results of other verification procedures used with audio recordings (Chen, 1999; Ferber, 1995).

Speech errors in natural conversations are collected by human listeners, and as such, they are subject to a set of perceptual biases that may skew the distributions of certain errors (Bock, 1996; Cutler, 1982; Pérez et al., 2007). The collection methods described above for SFUSED are demonstrably less prone to bias than prior studies that used different methods. First, the SFUSED methods result in very high sample coverage. Using capture-recapture methods from data collected by multiple listeners, we have calculated that a speech error of any kind occurs once every 48.5 seconds. A sample of a recently analyzed podcast shows that three collectors listening to the same recording collect a valid error, on average, once every 82.4 seconds, or they detect

about 59% of the actual errors. A comparable estimate from the Garnham et al. (1981) collection detects an error about once every 355.8 seconds, or about 14% of the data.

In addition to differences in sample coverage, it is relatively clear that the methods employed in building SFUSED results in fewer errors that are highly salient and therefore easy to detect. For example, exchanges, like *torn korkilla* (sfusedE-1495), are very salient and therefore usually not missed by human listeners (Stemberger, 1982/1985). Exchanges of all types (sound exchanges, word exchanges, etc.) account for about 0.38% of the total errors in SFUSED, but they have much higher percentages in other collections: between 5-6% in traditional online collections ((Boomer & Laver, 1968; Nooteboom, 1969; Stemberger, 1982/1985) and even as much as 35% (Pérez et al., 2007) and 54% (Dell & Reich, 1981) in collections that use large numbers of untrained collectors.

Because the SFUSED database is less skewed toward errors that are easy to hear, it produces a better sample of errors that are more difficult to hear. For example, prior research has shown that mis-pronunciations that involve a change of place of articulation are easier to detect than changes in voicing (Cole, Jakimik, & Cooper, 1978). Because SFUSED contains both errors collected online and offline, we can compare the rates of place-changing errors relative to voice-changing errors. It turns out that place-changing errors are about three times as frequent as voice-changing errors in the online data, but the order is reversed in the offline data, with voicing change errors being slightly more common. In sum, because of the improved sample coverage, the SFUSED database gives a more accurate sample of the actual occurrence of errors in natural speech.

In order to assess phonotactic violations, we require a system of phonotactics. Following standard practice (e.g., Giegerich (1992)), we describe phonotactics with reference to syllable structure. Our detailed assumptions about syllable structure and how restrictions on syllabic positions constitute phonotactic constraints are described in the appendix. To summarize briefly, we assume that most consonants can be simple syllable onsets (except η and ζ), but complex CC onsets are more restricted, essentially requiring an obstruent followed by a non-nasal sonorant. Nuclei can be either filled by tense or lax vowels, which leads to certain restrictions on word position and following coda consonants: tense vowels cannot occur with a complex CC coda, and lax vowels are not allowed in final open syllables. Complex codas must fall in sonority. An initial “appendix” s is allowed initially in all words, and a string of up to three coronal appendices are allowed word-finally. This basic description is supplemented in the appendix with a number of specific constraints on natural classes tied to particular positions. Our goal here is not to argue for a particular theory of English phonotactics, but rather to adopt a standard one and apply it consistently to our speech error data.

4.2 Results and discussion

We assess phonological regularity below by reporting percentages of phonotactic violations in different syllable contexts and speech error types. In particular, we examined a set of 2,228 sound errors and word blends and applied the phonotactic constraints described above and in the appendix to the surface form of the error word. The general finding is that irregularity (i.e., percentages of phonotactic violations) is far higher than prior research. Because of this, we also assess the likelihood that these violations could have arisen by chance and assess production models based on these findings.

To give a sense of how our phonotactic constraints are applied to the data, Table 1 illustrates the different kinds of phonotactic effects, including breaking them down by syllable position and whether the error word contains a non-native segment. Many of the examples below

violate constraints on licit onsets, either producing banned consonant clusters, e.g., /θl vr pw/ or, rarely, placing a sound that cannot be in onset position, as in [ʒ]u for ‘you’. While there are some medial illicit onsets, the majority of these violations (74%) occur in word-initial position. Illicit appendices involve either an initial extra consonant added to a valid onset that is not [s], or an extra consonant that is one of the following set of coronal obstruents [t d s z θ ð dʒ tʃ]. Cases like [ʃpr]eadsheet with an initial [ʃ] are very common: 7 out of the 10 illicit appendices involve this segment (see Stemberger (1983) for a similar pattern). Phonotactic violations also involve illicit codas, like clusters that do not fall in sonority or have licit appendices, e.g., spi[lkf], and illicit nucleus + coda combinations: a tense vowel + [ŋ] in l[aiŋ] for ‘lane’, or a tense vowel followed by a coda cluster, as in came → c[eɪmp]. Cases like NP → N[pʌ] violate the rule that lax vowels cannot be in final open syllables. Our sample also includes error words that contain sounds that are clearly outside English’s phonemic inventory, as in the front rounded vowel in the aborted attempt at ‘clearly’, cl[y]=.

Table 1. Phonotactic violations by syllable context

| Context and type | Examples ² | N |
|----------------------|--|-------------|
| Illicit onset | <i>listen</i> → [θl]isten, <i>viral</i> → [vr]iral, <i>things/ways</i> → [θw]ays, <i>Screetch</i> → [sr]eetch, <i>why</i> → [pw]y, <i>after</i> → af[ts]er, <i>you</i> → [ʒ]u, <i>means</i> → [mr]eans | 45 (44.12%) |
| Illicit appendix | <i>store</i> → [ʃt]ore, <i>screen</i> → [ʃkr]een, <i>spreadsheet</i> → [ʃpr]eadsheet, <i>special book</i> → [sb]ook, <i>works</i> → work[ʃ] | 10 (9.8%) |
| Illicit coda | <i>Christmas</i> → Chri[fs]mas, <i>speak film</i> → spi[lkf] | 5 (4.9%) |
| Illicit nucleus/rime | <i>lane</i> → l[aiŋ], <i>came</i> → c[eɪmp], <i>NP</i> → N[pʌ], <i>show</i> → sh[ʊ], <i>post-pubescent</i> → p[jou]= | 29 (28.43%) |
| Non-native segment | <i>week</i> → w[e]k, <i>best</i> → be[ɪ], <i>clearly</i> → cl[y]= | 13 (12.75%) |
| <i>Total</i> | | 102 |

Another way to explore how phonotactics can be violated is by examining the incidence of violations by speech error type. Table 2 below quantifies regularity as a percentage of regular and irregular forms within each error type. Regularity clearly differs by error type in ways that should be expected given the nature of English phonotactics. Thus, as a percentage of total, there are roughly six times more violations in additions than in deletions. This difference makes sense because the majority of English phonotactic constraints are restrictions on consonant combinations, and additions create them while deletions eliminate them. Likewise, the relatively high rate of violations in both blend types can be explained by the fact that blends merge segments from different word contexts, so the combinations are more random than the inputs to errors that work on a single word form.

² These examples can be explored in SFUSED with the following record IDs, as ordered in the table: 1236, 1278, 7120, 3954, 4187, 5545, 5739, 5599; 1224, 4278, 2105, 4453, 6448; 4398, 7211; 1245, 1526, 2223, 7348, 6844; 2811, 3035, 5964

Table 2. Regularity by error type

| | Regular ³ | | Irregular | |
|-------------------|----------------------|---|-------------|--------------------------------------|
| Substitutions | 1,416 (96.92%) | blood → b[r]ood | 45 (3.08%) | store → [ʃt]ore |
| Additions | 353 (89.59%) | thought → th[r]ought | 41 (10.41%) | viral → [vr]iral |
| Deletions | 181 (98.37%) | glow → go | 3 (1.63%) | Screech → [sr]eech |
| Exchanges | 35 (94.59%) | corn tortilla → [t]orn [k]ortilla | 2 (5.41%) | Moonrise Kingdon → Moonrai[ŋ] Ki[z]= |
| Shifts | 8 (100%) | spare blackforests → splare backforests | 0 | |
| Sequential blends | 60 (90.91%) | three fifty → thrifty | 6 (9.09%) | speak film → spi[lkf] |
| Word blends | 73 (93.59%) | trust/faith → traith | 5 (6.41%) | things/ways → [θw]ays |
| <i>Total</i> | 2126 (95.42%) | | 102 (4.58%) | |

The overall percentage of phonotactic violations (=irregularity) is roughly five times greater than the percentage given in Stemberger (1983). Given the differences in methods, and the ability to replay speech samples, it makes sense to ask if this difference is due to methodology. It turns out that SFUSED contains data appropriate for investigating this question. In the initial stage of developing our collection methods, the first author and his team collected roughly 1,100 speech errors of all kinds using an online method similar to Stemberger’s method. We can therefore test for an effect of collection method by distinguishing datasets collected online versus those collected offline. This comparison is made in Table 3. There are more online than offline errors because these datasets are balanced for experience level, and there is more data from online sources from the particular level we used. The online errors, at 1% irregularity, have roughly the same amount of phonotactic violations as Stemberger’s (1983) count, which also used an online method. The rate of irregularity in the offline errors is roughly three times greater, which is significant ($\chi(1)^2=7.902, p=0.0049$).

Table 3. Regularity and collection method

| | Offline | Online |
|--------------------------|--------------|--------------|
| Phonologically regular | 516 (96.81%) | 831 (99.05%) |
| Phonologically irregular | 17 (3.19%) | 8 (0.95%) |

These facts strongly suggest that the reported rate of phonological regularity is affected by methodology. Furthermore, different studies may employ different sets of phonotactic constraints when assessing regularity: Stemberger (1983) notes that Fromkin (1971) does not assume initial [ʃC] clusters like [ʃt]ore for ‘store’ are irregular. So it is possible that different assessments are explained by different phonotactic systems. However, this lack of consistency cannot account for the differences between online and offline regularity because the same system of phonotactics was used for both datasets. It seems clear, as conjectured by Cutler (1982) and Shattuck-Hufnagel (1983), that perceptual bias affects phonological regularity.

This conclusion suggests that the offline data provides the best gauge of regularity. Removing the online errors from the larger counts above yields a rate of 94.5% regular vs. 5.5%

³ The phonotactically regular examples can be explored in SFUSED with the following record IDs: 888, 142, 24, 1495, 1641, 885, 1471

irregular speech errors. While regularity is much lower than in prior research, these rates still support the contention that sub-lexical speech errors are overwhelmingly regular. However, they raise the issue of whether attested regularity is greater than it would be expected by chance. After all, speech errors are deviations from an established speech plan, and the words of that speech plan are phonologically well-formed. Perhaps a commitment to that plan alone, which is formalized in all models, could account for the fact of overwhelming regularity. To investigate this, we performed a set of permutation tests similar to that developed in Dell and Reich (1981), to estimate chance occurrence of phonologically regular words.

Our tests involve random permutations of segments from a list of intruder segments of a particular error type, holding constant the slot in the syllable for the permuted segment. For example, consider C1 substitutions into C1C2 complex onsets. The error corpus gives us a set of potential intruders in this slot. By randomly permuting the list of intruders and then inserting them into a list of slots, we obtain a sample of what the data would look like if the slots and intruders were independently selected for the errors. We perform multiple trials of this random permutation, from which we are able to obtain a distribution of the percentage of regular errors under the assumption of independence. Taking an average of the percentages over the distribution gives the expected percentage of how often this substitution produces phonotactic violations of this type.

As shown in Table 4, it turns out that random consonant substitutions into C2 were phonologically regular 78% of the time, which is less than what is observed in our data (81%). But the p -value (the probability that the random permutation yields a percentage of 81% or higher) is 0.38, which is not significant (see Kessler (2001) for an explanation of how to obtain p -values for a permutation test). Table 4 gives the results of these tests for substitutions into C1 and C2 positions in both substitutions and additions. Interestingly, there is clearly no evidence to reject the independence hypothesis in the two cases where C1 is the intruder in the C1C2 onset. On the other hand, there is strong evidence to reject the independence hypothesis when C2 is the intruder. For example, C2 substitutions (*dream* → *dweam*) yield no phonotactic violations at all in the entire corpus of errors. But the permutation test shows that plenty could have occurred (*trick* → *tlick*). So independence of C1 and C2 is strongly rejected.

Table 4. Actual and randomly generated regularity

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

What are the implications of these facts for models of phonological encoding? Perhaps the best way to relate these findings to models is to compare predicted regularity with the facts reported above. As discussed in the introduction, the simple recurrent network model of Dell et al. (1993) was developed specifically to investigate how a model of phonological encoding respects phonotactics. Different model parameters produced different results, but the best model (trained on frequent vocabulary, with internal and external representation of prior segments, and with correlated input representations) produced words that were phonologically regular 96.5% of the time. Other model parameters produced regularity in the 89-95% range. These modelling results compare well with the facts reported above, focusing on the data collected offline, which were regular 94.5% of the time. Many of the errors produced by the simple recurrent network

model indeed resemble the errors we have found in SFUSED, dominated by illegal clusters and initial sequences. The simple recurrent network model also predicts a larger number of errors in word-initial position, a well-attested pattern in many languages (Wilshire, 1998), and this prediction accords with our analysis of the difference between regularity in both substitutions and additions in C1 of a CC onset. Some limitations prevent us from concluding our results are a direct match with this model's predictions, including the fact that Dell et al. only focused on monosyllabic words, and their model actually did not allow for additions. However, the larger results are clearly consistent with the model's predictions and future research could investigate these details.

5. General discussion

We have explored different dimensions of phonological regularity and how they may shape speech errors, in particular the sound structure of sub-lexical errors. In addition to global assessments of the sound structure of words, classified categorically as regular or irregular, we have described regularity at the granular level as well. In a sense, phonological regularity is constituted by conformity to a host of specific measures on different dimensions of analysis: markedness measures, frequency measures, and even measures for default segments that run counter to these expectations. Indeed, different models of phonological encoding make rather different predictions of the sound structure of speech errors within these dimensions.

We have documented phonological regularity in the SFUSED speech error corpus and found that the standard of phonological regularity is much lower than suggested by prior research. In particular, this evidence suggests that a 99% standard of phonological regularity from Stemberger (1983) is too high, and the actual standard, which depends on error type and position within a word, is closer to 94.5%. It is rather clear that this difference in observed phonotactic violations is due to perceptual biases, because SFUSED is demonstrably less prone to perceptual bias than other models that use offline data collection. This finding has implications for language production models because it provides a reasonably good fit with a model that has been shown to make predictions about phonological regularity through computational simulation, namely the simple recurrent network model of Dell et al. (1993). Other models can also be implemented to test predictions about phonological regularity and tested against these results.

5.1 Perceptual biases and regularity

The conclusion that phonological regularity is affected by methodology leads to new empirical questions about the scope of grammatical regularity. Because prior work has focused largely on global phonological regularity, we have focused our investigation in section 4 on this categorical distinction. But perceptual bias could also have an impact on the shape of phonological structures that underlie this categorical assessment. For example, research on the perception of errors has demonstrated that certain phonological structures are detected more readily than others, including differences in place class, manner, and voicing categories (Cole et al., 1978; Garnes & Bond, 1975; Marslen-Wilson & Welsh, 1978). Clearly, these distinctions are relevant to characterization of the granular regularity for markedness and frequency, and so the error counts in prior studies involving these distinctions may also differ from what we find in SFUSED.

At the word level, there is reason to believe that word-level errors may also be affected by perceptual bias. Consider the category constraint, according to which intended and error words are of the same part of speech, and so it is generally syntactically regular. Estimates of the

respect for this constraint range from between 100% (Nootboom, 1969) and 85% (Garrett, 1980). It seems plausible that category constraint errors, because of their syntactic ill-formedness, might also be regularized or excluded as false starts, and that closer attention to this kind of error through audio recordings might reveal weaker support for the constraint itself. Indeed, a sample of 760 lexical substitutions from a larger set of 4,003 speech errors lends some support to this conjecture. In this sample, which included substitutions of both lexical and functional items, 640 errors respected the category constraint, which, at 84.5%, is slightly below the lowest assessment from prior research.

Sentence blends and word shifts also support the contention that, while syntactic errors tend to be regular, they can on occasion go awry. Sentence blends merge two phrases that communicate roughly the same message. They often result in semantically bizarre but syntactically well-formed sentences, as in ... *at the same sense* (sfusedE-1161, blend of *at the same time* and *in the same sense*). However, they also often result in strongly ungrammatical sentences, as illustrated by *I was just had a thought* (sfusedE-5755, merging of *I was just thinking* and *I just had a thought*). Word shifts, while rare, present a more drastic situation syntactically. In a sample of 27 such errors, nine were found to be syntactically well-formed, two marginal, and 16 strongly ungrammatical, as in *I'm not staying all up night* (sfusedE-1817). These syntactic errors are rare, but only about a third of the word shifts in SFUSED are syntactically regular. Thus, as with sound errors, different types of errors have different proclivities toward regularity, there does seem to be clear evidence for a stronger trend of grammatical ill-formedness than previously acknowledged in the literature.

One model that may provide the basis to formalized predictions about syntactic well-formedness is the dual path model (Chang, 2002; Chang, Dell, & Bock, 2006). This model is a full-fledged model of sentence production that combines a word sequencing system with a meaning system for parsing pre-linguistic messages. This model employs a simple recurrent network with hidden layer representations to map prior words to the next word in a sequence. In a sense, it is like Dell et al.'s (1993) simple recurrent network because it works with a similar underlying architecture and predicts sentence structure without the explicit use of sentence frames. The model learns from listening, or deviations from heard sentences and sentences expected by the model, which lead to weight changes that gradually lead to better approximations of natural speech. We believe that this would be a good model for testing predictions about syntactic regularity because it appears to have a robust capacity for encoding syntactic regularity in the sequencing system, and, at the same time, potential for producing syntactically ill-formed sentences, as attested occasionally in natural speech.

5.2 Markedness, frequency, and anti-frequency

As laid out empirically (section 2) and theoretically (section 3), there are rather different perspectives of the impact of granular regularity in speech errors. On the one hand, we have competing explanations for the over-abundance of some phonological structures in errors. Both markedness theory and psycholinguistic theories that explicitly encode type frequency predict that frequent/unmarked structures should be more prevalent in speech errors. On the other hand, the so-called anti-frequency effects, like the palatal bias, introduce a conflicting empirical account at the granular level. For example, Levitt and Healy (1985) and Stemberger (1991a) have exactly opposite predictions about the role of frequency in experimentally elicited speech errors. It seems, therefore, that progress can be made by enriching the empirical basis for teasing apart these different effects.

The conflicting solutions of frequency and anti-frequency seem like a more tractable problem. First, prior research has uncovered some problems with the reported anti-frequency, including perceptual biases that could account for the increased counts of low frequency sounds (Marin, Pouplier, & Harrington, 2010; Pouplier & Goldstein, 2005), frequency effects that emerge in short term implicit priming experiments (Goldrick, 2011), and confounds with the distinction between lexical and non-lexical words (Frisch, 1996). Another empirical issue that strikes us as important is that the statistical methods employed by studies arguing for certain anti-frequency biases (Shattuck-Hufnagel & Klatt, 1979; Stemberger, 1991a) do not seem to have related differences in error occurrence to baseline frequencies in robust ways. Another issue is that conflicting accounts of (anti-)frequency in experimentally elicited errors have used rather different tasks (Frisch, 1996; Levitt & Healy, 1985; Stemberger, 1991a). Therefore, more direct tests using the same types of stimuli and experimental tasks may resolve some of these issues. A final point is that frequency and anti-frequency make very different predictions in so many contexts of the sound system, empirical confirmation in one direction or the other should be rather straightforward. Perhaps when languages other than English are examined, and methods less prone to perceptual bias are employed, the larger picture will be more definitive.

The situation is less straightforward for the competing solutions with markedness and frequency because the two approaches tend to have overlapping predictions for most of the sound structures of English. For example, the stimuli employed in Goldrick (2002) employed consonant pairs like /t s/ that reveal differences between the predictions of markedness (favors /t/) and frequency (favors /s/), but there are only a small handful of consonant pairs in English where such comparisons are possible. Another limitation of our current understanding is that the evidence for phonological markedness is dominated by research on the speech of aphasics. We believe that future research can use the findings from this research to inform designs of parallel empirical questions about speech in normal populations. Finally, given the fact that markedness and frequency give competing explanations for so much of the data in English, a look to languages with different sound structure from English has significant potential for filling out the empirical picture.

As a final note, we worry that we may be falling into the trap of reductionism. It is rather common that scientific pursuits attempt to associate a particular problem with a single unifying solution. However, considering the theoretical perspectives laid out above, there is no obvious underlying principle that would prevent combining explanations offered by markedness and frequency, and even anti-frequency if it is truly attested. Complex human behaviors like speech often have many factors influencing observed behavior. Indeed, some researchers have explicitly argued for more complex solutions involving both markedness and frequency (Buchwald, 2005; Goldrick & Larson, 2010). Some of the models outlined in section 3 also provide a theoretical framework for formalizing this complexity. For example, models based on Harmonic Grammar propose that error outcomes are shaped by markedness, but the impact of these constraints is scaled by constraint weights. In standard error-corrective learning, the weights are learned through exposure to natural data, which is of course sensitive to the frequency of sound structure. Another way to explore this problem, therefore, is to determine how learning could lead to the appropriate weight assignments, and if the production systems based on those assignments have a good fit with the error data.

Appendix

All segments in the examples have syllabic roles, e.g., onset consonants, nucleus vowels, etc. Our syllabification assumptions are rather standard and closely follow established formal accounts of English syllabification (Giegerich, 1992; Jensen, 1993; Selkirk, 1982). We assume that all words may have an initial appendix, which must be [s] initially, but has more possibilities finally (see below). A syllable is composed of an obligatory nucleus, which is either a monophthong or diphthong, and may also have a [j] onglide. Syllables may optionally have onsets or codas, and the possible options for English are shown below. Finally, the rime is a unit that combines the nucleus and the coda. Our syllable structure is surface oriented and does not assume “ambisyllabic” consonants, though one can easily modify our assumptions to posit ambisyllabic consonants if needed.

Onsets: (s) (C₁)(C₂)

General:

- All C positions are optional
- A C₁ on its own can be any consonant except *ŋ ʒ* (*ʒ* okay word-internally)
- Appendix (*s*) is only allowed initially (also exceptionally *f* in certain words)

(C₁)(C₂) clusters:

- C₁ = obstruent + C₂ = non-nasal sonorant (so assume *sm* has an appendix)
- No voiced fricative + sonorant, i.e., **v ð z ʒ + l r w*
- No affricate + sonorant, i.e., **tʃ dʒ + l r w*
- No coronal non-strident obstruent + lateral, i.e., **t d θ + l*
- No labial + *w*, i.e., **p b f + w*
- Also these combinations banned: *fw fl sr sh gw θw fjV*

Appendix + Onset:

- In *s + C₁*, C₁ is a voiceless stop, so: *s + p t k*, assuming *sf* are rare/loan material
- In *s + C₁C₂*, **stw*, **skl*

Assumptions about onglide *j*

- Onglide *j* is not governed by the C₂ phonotactics above, but instead is considered part of the peak, because it only combines with the tense vowel *u*, as in *few* [fju]
- However, there are no *CCju* words, so in a sense *j* fills the C₂ position, because it pre-empts a complex onset

Codas: (C₁)(C₂) (C₃)(C₄)(C₅)

General

- A single coda C can be any consonant, except *h*, and *j w* are generally assumed to be consonantal in onset position and vocalic in peak position, which overlaps in a sense with C₁ in some analyses because it is the second part of the peak
- All consonants are optional
- C₁ is always a sonorant
- C₃₋₅ are final appendices and can only be coronal obstruents, i.e., *t d s z θ ð dʒ tʃ*
- ŋ* is only allowed in coda position of a syllable containing a lax vowel (see below)

C₁C₂ clusters:

- In one possible C₁C₂ cluster, C₂ falls in sonority from C₁ on the following scale: *r > l > m n ŋ > obstruents*; thus *Vrl* is okay, but **Vlr*
- Another option is *s + voiceless stop p t k*, e.g., *bask, mast*

- c. Nasal + obstruent clusters must agree in primary place (*mp ŋk*, **mk *ŋp*) and the obstruent must be voiceless (**mb*, **mg*)
- d. Also, the following are banned codas: **lg*

Nucleus: X_1 (X_2) (a.k.a ‘Peak’)

- a. The first slot X_1 is always a vowel; words like *butter* that have syllabic sonorants in some analyses are analyzed [tʌr], with a schwa + coda C. X_2 is either part of a vowel or overlaps with C_1 from the coda.
- b. Tense vowels *i (a) u* and diphthongs *eɪ oʊ oɪ aɪ aʊ* are considered bimoraic and so occupy two nucleus slots, while lax vowels *ɪ ɛ (æ) ʊ ʌ ə* are monomoraic and only occupy one X slot. No special length diacritics are used for tense vowels.
- c. High vocoids /i j/ and /u w/ are considered the same phoneme, vocalic in Nucleus position, and transcribed as such, and consonantal in onglide or onset positions. They are banned from coda positions.
- d. Stressed syllables and final syllables must have X_1+X_2 , i.e., be bimoraic.
- e. Syllables are maximally trimoraic, where X_1+X_2 + one coda C have moras.
- f. These assumptions mean that tense vowels can only be in VV(C) rimes (ignoring appendices) in stressed syllables, because they are inherently bimoraic.
- g. These assumptions mean that lax vowels can only be in VC(C) rimes in stressed and final syllables, because they are inherently monomoraic and therefore only occupy one X slot.
- h. However, lax vowels do occur in many non-final stressed open syllables, and in this case, the typical assumption is that the following onset is ambisyllabic in order to fill the X_2 slot and satisfy the bimoraic requirement.

References

- Archangeli, D. (1984). *Underspecification in Yawelmani Phonology and Morphology*. (Doctoral dissertation), Massachusetts Institute of Technology.
- Béland, R. (1990). Vowel epenthesis in aphasia *Morphology, phonology, and aphasia* (pp. 235-252). New York: Springer-Verlag.
- Béland, R., & Paradis, C. (1997). Principled syllabic dissolution in a primary progressive aphasia case. *Aphasiology, 11*, 1171-1196.
- Béland, R., Paradis, C., & Bois, M. (1993). Constraints and repairs in aphasic speech: A group study. *Canadian Journal of Linguistics, 38*, 279-302.
- Blumenstein, S. (1973). *A phonological investigation of aphasic speech*. The Hague: Mouton.
- Bock, K. (1982). Toward a cognitive psychology of syntax: information processing contributions to sentence formulation. *Psychology Review, 89*, 1-47.
- Bock, K. (1996). Language production: Methods and methodologies. *Psychonomic Bulletin & Review, 3*, 395-421.
- Boersma, P., & Pater, J. (2008). *Convergence properties of the gradual learning algorithm for Harmonic Grammar*. Ms., University of Amsterdam and University of Massachusetts, Amherst.
- Bond, Z. S. (1999). *Slips of the ear: Errors in the perception of casual conversation*. San Diego: Academic Press.
- Boomer, D. S., & Laver, J. D. M. (1968). Slips of the tongue. *International Journal of Language and Communication Disorders, 3*, 2-12.
- Broeke, V. D. M. P. R., & Goldstein, L. (1980). Consonant features in speech errors. In V. Fromkin (Ed.), *Errors in linguistic performance: Slips of the tongue, ear, pen, and hand* (pp. 47-65). London: Academic Press.
- Buchwald, A. (2005). *Representation, repair, and well-formedness: Grammar in spoken language production*. (Doctoral dissertation), Johns Hopkins University.
- Buchwald, A. (2009). Minimizing and optimizing structure in phonology: Evidence from aphasia. *Lingua, 119*, 1380-1395.
- Buchwald, A., Rapp, B., & Stone, M. (2007). Insertion of discrete phonological units: An ultrasound investigation of aphasic speech. *Language and Cognitive Processes, 22*, 910-948.
- Chang, F. (2002). Symbolically speaking: A connectionist model of sentence production. *Cognitive Science, 26*, 609-651.
- Chang, F., Dell, G. S., & Bock, K. (2006). Becoming syntactic. *Psychological Review, 113*, 234-272.
- Chen, J.-Y. (1999). The representation and processing of tone in Mandarin Chinese: Evidence from slips of the tongue. *Applied Psycholinguistics, 20*, 289-301.
- Cole, R. A., Jakimik, J., & Cooper, W. E. (1978). Perceptibility of phonetic features in fluent speech. *Journal of the Acoustical Society of America, 64*, 45-56.
- Cutler, A. (1982). The reliability of speech error data. In A. Cutler (Ed.), *Slips of the tongue and language production* (pp. 7-28). Berlin: Mouton.
- Dell, G. S. (1986). A spreading-activation theory of retrieval in sentence production. *Psychological Review, 93*, 283-321.
- Dell, G. S. (1988). The retrieval of phonological forms in production: Tests of predictions from a connectionist model. *Journal of Memory and Language, 27*, 124-142.

- Dell, G. S. (1995). Speaking and misspeaking. In L. R. Gleitman & M. Liberman (Eds.), *An invitation to cognitive science, Language, Volume 1*. Cambridge, MA: The MIT Press.
- Dell, G. S., Juliano, C., & Govindjee, A. (1993). Structure and content in language production: A theory of frame constraints in phonological speech errors. *Cognitive Science*, *17*, 149-195.
- Dell, G. S., Reed, K. D., Adams, D. R., & Meyer, A., S. (2000). Speech errors, phonotactic constraints, and implicit learning: A study of the role of experience in language production. *Journal of Experimental Psychology: Learning, Memory and Cognition*, *26*, 1355-1367.
- Dell, G. S., & Reich, P. A. (1981). Stages in sentence production: An analysis of speech error data. *Journal of Verbal Learning and Verbal Behavior*, *20*, 611-629.
- Elman, J. (1990). Finding structure in time. *Cognitive Science*, *14*, 179-211.
- Fay, D., & Cutler, A. (1977). Malapropisms and the structure of the mental lexicon. *Linguistic Inquiry*, *8*, 505-520.
- Ferber, R. (1995). Reliability and validity of slip-of-the-tongue corpora: A methodological note. *Linguistics*, *33*, 1169-1190.
- Frisch, S. (1996). *Similarity and frequency in phonology*. (Doctoral dissertation), Northwestern University.
- Frisch, S., & Wright, R. (2002). The phonetics of phonological speech errors: An acoustic analysis of slips of the tongue. *Journal of Phonetics*, *30*, 139-162.
- Fromkin, V. (1971). The non-anomalous nature of anomalous utterances. *Language*, *47*, 27-52.
- Garnes, S., & Bond, Z. S. (1975). Slips of the ear: Errors in perception of casual speech. *Proceedings of the 11th regional meeting of the Chicago Linguistics Society* (pp. 214-225).
- Garnham, A., Shillcock, R. C., Brown, G. D., Mill, A. I., & Cutler, A. (1981). Slips of the tongue in the London-Lund corpus of spontaneous conversation. *Linguistics*, *19*, 805-818.
- Garrett, M. (1975). The analysis of sentence production. In G. H. Bower (Ed.), *The psychology of learning and motivation, Advances in research and theory, vol. 9* (pp. 131-177). New York: Academic Press.
- Garrett, M. (1980). The limits of accommodation: Arguments for the independent processing levels in sentence production. In V. Fromkin (Ed.), *Errors in linguistic performance: Slips of the tongue, ear, pen, and hand* (pp. 263-271). New York: Academic Press.
- Giegerich, H. J. (1992). *English phonology: An introduction*. Cambridge: Cambridge University Press.
- Goldrick, M. (2002). *Patterns of sound, patterns in mind: phonological regularities in speech production*. (Doctoral dissertation), Johns Hopkins University.
- Goldrick, M. (2004). Phonological features and phonotactic constraints in speech production. *Journal of Memory and Language*, *51*, 586-603.
- Goldrick, M. (2011). Linking speech errors and generative phonological theory. *Language and Linguistics Compass*, *5*(6), 397-412.
- Goldrick, M., & Daland, R. (2009). Linking speech errors and phonological grammars: Insights from Harmonic Grammar networks. *Phonology*, *26*, 147-185.
- Goldrick, M., & Larson, M. (2008). Phonotactic probability influences speech production. *Cognition*, *107*, 1155-1164.

- Goldrick, M., & Larson, M. (2010). Constraints on acquisition of variation. In C. Fougeron, B. Kühnert, M. D'Imperio, & N. Vallée (Eds.), *Laboratory phonology 10: Variation, phonetic detail, and phonological representation*. Berlin: Mouton de Gruyter.
- Goldrick, M., & Rapp, B. (2007). Lexical and post-lexical phonological representations in spoken production. *Cognition*, 102, 219-260.
- Greenberg, J. (1966). Language universals. In T. Sebeok (Ed.), *Current Trends in Linguistics* (Vol. 3). The Hague: Mouton.
- Hayes, B., & Wilson, C. (2008). A maximum entropy model of phonotactics and phonotactic learning. *Linguistic Inquiry*, 39, 379-440.
- Hockett, C. F. (1967). Where the tongue slips, there slip I *To Honor Roman Jakobson, Vol II*. The Hague: Mouton.
- Itô, J. (1986). *Syllable theory in prosodic phonology*. (Doctoral dissertation), University of Massachusetts, Amherst.
- Jakobson, R. (1941). *Child language, aphasia and phonological universals*. The Hague: Mouton.
- Jensen, J. T. (1993). *English phonology* (Vol. v. 99). Amsterdam ; Philadelphia: J. Benjamins Pub. Co.
- Jordan, M. I. (1986). Attractor dynamics and parallelism in a connectionist sequential machine *Proceedings of the Eight Annual Conference of the Cognitive Science Society* (pp. 531-546). Hillsdale, NJ: Lawrence Erlbaum.
- Kessler, B. (2001). *The significance of word lists*. Stanford: CSLI Publications.
- Kessler, B., & Treiman, R. (1997). Syllable structure and the distribution of phonemes in English syllables. *Journal of Memory and Language*, 37, 295-311.
- Kiparsky, P. (1982). From cyclic phonology to lexical phonology. In H. v. d. Hulst & N. Smith (Eds.), *The structure of phonological representations* (Vol. 1, pp. 131-175). Dordrecht: Foris.
- Kupin, J. J. (1982). *Tongue twisters as a source of information about speech production*. Bloomington: Indiana University Linguistics Club.
- Lahiri, A., & Marslen-Wilson, W. D. (1992). Lexical processing and phonological representation. In D. R. Ladd & G. J. Docherty (Eds.), *Papers in Laboratory Phonology: Gesture, Segment, Prosody* (pp. 229-254). Cambridge: Cambridge University Press.
- Legendre, G., Miyata, Y., & Smolensky, P. (1990). Can connectionism contribute to syntax? Harmonic Grammar, with an application. In M. Ziolkowski, M. Noske, & K. Deaton (Eds.), *Proceedings of the 26th Regional Meeting of the Chicago Linguistic Society* (pp. 237-252). Chicago: Chicago Linguistic Society.
- Levelt, W. J. M., Roelofs, A., & Meyer, A., S. (1999). A theory of lexical access in speech production. *Behavioral and Brain Sciences*, 22, 1-75.
- Levin, J. (1985). *A Metrical Theory of Syllabicity*. (Doctoral dissertation), MIT, Cambridge, Massachusetts.
- Levitt, A., & Healy, A. (1985). The roles of phoneme frequency, similarity, and availability in the experimental elicitation of speech errors. *Journal of Memory and Language*, 24, 717-733.
- MacKay, D. G. (1970). Spoonerisms: The structure of errors in the serial order of speech. *Neuropsychologia*, 8, 323-350.
- Marin, S., Pouplier, M., & Harrington, J. (2010). Acoustic consequences of articulatory variability during productions of /t/ and /k/ and its implications for speech error research. *The Journal of the Acoustical Society of America*, 127(1), 445-461.

- Marslen-Wilson, W., & Welsh, A. (1978). Processing interactions and lexical access during word recognition in continuous speech. *Cognitive Psychology*, *10*, 29-63.
- Meyer, A. S., & Belke, E. (2007). Word form retrieval in language production. In G. Gaskell (Ed.), *The Oxford handbook of psycholinguistics* (pp. 471-487). New York: Oxford University Press.
- Motley, M. T., & Baars, B. J. (1975). Encoding sensitivities to phonological markedness and transitional probability: Evidence from spoonerisms. *Human Communication Research*, *1*, 353-361.
- Nooteboom, S. G. (1969). The tongue slips into patterns. In A. J. van Essen & A. A. van Raad (Eds.), *Leyden studies in linguistics and phonetics* (pp. 114-132). The Hague: Mouton.
- Paradis, C. (1988). On constraints and repair strategies. *The Linguistic Review*, *6*, 71-97.
- Paradis, C., & LaCharité, D. (Eds.). (1993). *Constraint-based Theories in Multilinear Phonology*.
- Paradis, C., & Prunet, J.-F. (Eds.). (1991). *The Special status of coronals: Internal and external evidence*. San Diego: Academic Press.
- Pérez, E., Santiago, J., Palma, A., & O'Seaghdha, P. G. (2007). Perceptual bias in speech error data collection: Insights from Spanish speech errors. *Journal of psycholinguistic research*, *36*, 207-235.
- Poupier, M., & Goldstein, L. (2005). Asymmetries in the perception of speech production errors. *Journal of Phonetics*, *33*, 47-75.
- Prince, A. (2007). The pursuit of theory. In P. de Lacy (Ed.), *The Cambridge handbook of phonology* (pp. 33-60): Cambridge University Press.
- Prince, A., & Smolensky, P. (1993/2004). *Optimality theory: Constraint interaction in generative grammar*. Malden, MA: Blackwell.
- Roelofs, A. (1992). A spreading-activation theory of lemma retrieval in speaking. *Cognition*, *42*, 107-142.
- Romani, C., & Calabrese, A. (1998). Syllabic constraints in the phonological errors of an aphasic patient. *Brain and Language*, *64*, 83-121.
- Romani, C., Olson, A., Semenza, C., & Granà, A. (2002). Patterns of phonological errors as a function of a phonological versus articulatory locus of impairment. *Cortex*, *38*, 541-567.
- Selkirk, E. (1982). Syllables. In H. v. d. Hulst & N. Smith (Eds.), *The Structure of Phonological Representations* (Vol. 2, pp. 337-383). Dordrecht: Foris.
- Shattuck-Hufnagel, S. (1979). Speech errors as evidence for a serial-ordering mechanism in sentence production. In W. E. Copper & E. C. T. Walker (Eds.), *Sentence processing: Psycholinguistic studies presented to Merrill Garrett* (pp. 295-342). Hillsdale, NJ: Erlbaum.
- Shattuck-Hufnagel, S. (1983). Sublexical units and suprasegmental structure in speech production planning. In P. F. MacNeilage (Ed.), *The production of speech* (pp. 109-136). New York: Springer Verlag.
- Shattuck-Hufnagel, S. (1986). The representation of phonological information during speech production planning: Evidence from vowel errors in spontaneous speech. *Phonology Yearbook*, *3*, 117-149.
- Shattuck-Hufnagel, S., & Klatt, D. H. (1979). The limited use of distinctive features and markedness in speech production: Evidence from speech error data. *Journal of Verbal Learning and Verbal Behavior*, *18*, 41-55.

- Smolensky, P., Goldrick, M., & Mathis, D. (2014). Optimization and quantization in gradient symbol systems: A framework for integrating the continuous and the discrete in cognition. *Cognitive Science*, *38*, 1107-1138.
- Stemberger, J. P. (1982/1985). *The lexicon in a model of language production*. New York: Garland.
- Stemberger, J. P. (1983). *Speech errors and theoretical phonology: A review*. Bloomington: Indiana University Linguistics Club.
- Stemberger, J. P. (1991a). Apparent antifrequency effects in language production: The addition bias and phonological underspecification. *Journal of Memory and Language*, *30*, 161-185.
- Stemberger, J. P. (1991b). Radical underspecification in language production. *Phonology*, *8*, 73-112.
- Stemberger, J. P. (1992). Vocalic underspecification in language production. *Language*, *68*, 492-524.
- Stemberger, J. P. (1993). Spontaneous and evoked slips of the tongue. In G. Blanken, J. Dittmann, H. Grimm, J. C. Marshall, & C.-W. Wallesch (Eds.), *Linguistic disorders and pathologies. An international handbook*. Berlin: Walter de Gruyter.
- Stemberger, J. P. (2004). Neighbourhood effects on error rates in speech production. *Brain and Language*, *90*, 413-422.
- Stemberger, J. P., & Treiman, R. (1986). The internal structure of word-initial consonant clusters. *Journal of Memory and Language*, *25*, 163-180.
- Taylor, C. F., & Houghton, G. (2005). Learning artificial phonotactic constraints: Time course, durability, and relationship to natural constraints. *Journal of Experimental Psychology: Learning, Memory and Cognition*, *31*, 1398-1416.
- Trubetzkoy, N. (1939). *Grundzüge der phonologie*. Göttingen: Vandenhoeck and Ruprecht.
- Vitevitch, M. S. (2002). Naturalistic and experimental analyses of word frequency and neighborhood density effects in slips of ear. *Language and Speech*, *45*, 407-434.
- Vousden, J. I., Brown, G. D. A., & Harley, T. A. (2000). Serial control of phonology in speech production: A hierarchical model. *Cognitive Psychology*, *41*, 101-175.
- Wan, I.-P., & Jaeger, J. J. (1998). Speech errors and the representation of tone in Mandarin Chinese. *Phonology*, *15*(417-461).
- Warker, J., & Dell, G. S. (2006). Speech errors reflect newly learned phonotactic constraints. *Journal of Experimental Psychology: Learning, Memory and Cognition*, *32*, 387-398.
- Warker, J. A. (2013). Investigating the retention and time course of phonotactic constraint learning from production experience. *Journal of Experimental Psychology: Learning, Memory and Cognition*, *39*, 96-109.
- Wayment, A. (2009). *Assimilation as attraction: Computing distance, similarity, and locality in phonology*. (Doctoral), Johns Hopkins University.
- Wells, R. (1951). Predicting slips of the tongue. *Yale Scientific Magazine*, *3*, 9-30.
- Wilshire, C. E. (1998). Serial order in phonological encoding: an exploration of the 'word onset effect' using laboratory-induced errors. *Cognition*, *68*, 143-166.
- Wilshire, C. E. (1999). The "tongue twister" paradigm as a technique for studying phonological encoding. *Language and Speech*, *42*, 57-82.