Received: 25 March 2022 | Revised: 15 June 2022 | Accepted: 20 June 2022

DOI: 10.1111/lnc3.12469

REVIEW ARTICLE

WILEY

Language generality in phonological encoding: Moving beyond Indo-European languages

John Alderete[1] | **Padraig G. O'Séaghdha**[2]

[1]Simon Fraser University, Burnaby, British Columbia, Canada

[2]Lehigh University, Bethlehem, Pennsylvania, USA

Correspondence
John Alderete, Simon Fraser University, 8888 University Drive, Burnaby, BC V5A 1S6, Canada.
Email: alderete@sfu.ca

Funding information
Social Sciences and Humanities Research Council of Canada, Grant/Award Number: 435-2020-0193

Abstract

Theories of phonological encoding are centred on the selection and activation of phonological segments, and how these segments are organised in word and syllable structures in online processes of speech planning. The focus on segments, however, is due to an over-weighting of evidence from Indo-European languages, because languages outside this family exhibit strikingly different behaviour and require the processing of additional phonological structures. We review evidence from speech error patterns, priming and form encoding studies, and re-syllabification in several non-Indo-European languages, including Mandarin, Cantonese, Japanese, Arabic, Hindi, Korean, Thai, and Vietnamese. We argue that these languages deepen our understanding of the nature of phonological encoding because they require recognising language-particular differences in: the first selectable (proximate) units of phonological encoding, the phonological units processed as word beginnings, the dynamics of syllable emergence during encoding, and the varied manifestations of re-syllabification. A satisfactory and general account of phonological encoding must incorporate these rich phenomena.

1 | INTRODUCTION

Modern psycholinguistics views speech production not as a monolithic process of linking thoughts to fluid speech, but rather as a chain of distinct production processes at different levels of encoding (Bock & Ferreira, 2014; Bock & Levelt, 1994; Dell, 1986; Garrett, 1980; Harley, 1984; Rapp & Goldrick, 2000). For example, the words that articulate our thoughts result from a process of lexical selection that maps concepts to lemmas, which are morpho-syntactic words devoid of form. Lexical selection is itself part of a larger process of grammatical encoding that situates lemmas within larger structures such as sentences. The process of phonological encoding works on the results of these processes to retrieve the phonological structure of planned words and phrases. For example, the intention to speak the English word *cat* involves selecting the lemma *CAT* from a set of related candidate lemmas (e.g., *DOG*, *CAT*, *RAT*). To implement this intention, phonological encoding retrieves the serially-ordered segments for this lemma, $[k]_{Onset}$, $[æ]_{Nucleus}$, $[t]_{Coda}$, and assembles them into a syllabified representation.

A number of psycholinguistic models have been proposed to account for the facts of phonological encoding (Dell, 1986; Goldrick, 2002; Goldstein et al., 2007; Harley, 1984; Levelt et al., 1999; Shattuck-Hufnagel, 1979; Smolensky et al., 2014; Stemberger, 1982/1985; Vousden et al., 2000). These models sometimes take up different theoretical positions on issues, such as whether language production processes at different levels are encapsulated or interactive (Dell & O'Séaghdha, 1991; Levelt et al., 1991; Rapp & Goldrick, 2000). However, these debates can sometimes obscure the extensive common ground across models. First, these models implement phonological encoding within an activation-dynamics in which the intention to produce a lemma leads to the activation of the sounds that make up that lemma. The activation of the intended sounds usually leads to the selection and production of these sounds, but they may also be mis-selected, generating speech errors. Second, these models are generally time-based in the sense that the activation dynamics for selecting sounds is affected by speech rate, and different time intervals for speaking the same utterance may have different outcomes (Browman & Goldstein, 1992; Dell, 1995). Third, there is general agreement that phonological encoding is a process distinct from other language production processes, including the later process of phonetic encoding, which implements discrete phonological structures as articulatory actions (Browman & Goldstein, 1992; Laganaro, 2019).

In addition to these general theoretical assumptions, there exists a set of shared assumptions about the nature of phonological encoding itself that has come to typify the research. One might call this the 'consensus model' of phonological encoding in the sense that research to date has led to a consensus supporting these assumptions, though many also acknowledge some of its limits, including its lack of cross-linguistic breadth. First, phonological encoding is fundamentally a matter of activating and selecting phonological segments. Phonological features (like [+voice] for voiced sounds) and syllable structure may have an impact on the processing of segments, but they are not explicitly selected in phonological encoding. Second, segments in word onsets are privileged over other segments because these 'starting points' provide the entry point to encoding. Finally, phonological encoding has a set of mechanisms for aligning selected segments with the larger prosodic structures in which they occur. For example, this alignment mechanism may involve slotting selected segments into canonical positions in a syllable frame, like inserting [k] in the Onset position for the word *cat*. Prosodic alignment of segments can be viewed as rooted in lexical representations (e.g., Shattuck-Hufnagel, 1979), or alternatively, as emergent during phonological encoding (Levelt et al., 1999).

These core assumptions of the consensus model are empirically grounded, and as we show below, they reflect true generalisations in the languages investigated. At the same time, however, many have pointed to their limited cross-linguistic scope. As has been noted in anthologised volumes on word

form encoding in particular (Meyer & Belke, 2007), and language production in general (Costa et al., 2007; Griffin & Crew, 2012), most of the generalisations supporting the consensus model have come from investigations of Indo-European languages, and as such they do not fully reflect the diversity of human language. Indeed, literature surveys of language production research (Alderete, 2022; Chen et al., 2002) have shown that this research has a strong bias towards the Indo-European languages and other major languages of the world, reflecting the same bias found in the larger field of psycholinguistics (Anand et al., 2011).

This lack of linguistic diversity leads to two problems. First, we cannot have as much confidence as we would like for a theoretical enterprise that has been developed for a small set of structurally similar languages. Second, we simply do not know what we do not know about languages beyond Indo-European, and we are missing an opportunity for new discoveries and insights about the mechanisms that underlie language production processes. The primary aim of this article is to update the consensus view of phonological encoding sketched above, and explain how recent research with a broader cross-linguistic perspective, including work on typologically distinct languages such as Chinese languages, Japanese, Hindi, Thai, Korean, Vietnamese and Arabic, can enrich our understanding of the process. This wider cross-linguistic perspective has led us to re-examine the three core assumptions, namely the singular role of segments, the privileged status of starting points, and the integration of prosodic with selected segmental content.

2 | THE FUNDAMENTAL ROLE OF PHONOLOGICAL SEGMENTS

As introduced above, phonological encoding has generally been assumed to be fundamentally a process that activates, selects, and positions phonological segments. Several different kinds of evidence, albeit with limited cross-linguistic scope, support this view. First, the sub-lexical structure of words may involve reductions, neutralisations, deletions and additions of phonological segments in connected speech. For example, Cruttenden (2014) lists several dozen casual speech phonological rules in English, such as schwa absorption with re-syllabification in words like /səpɔrt/ → [spɔrt] 'support' (see Bauer and Benedict (1997) for casual speech phonology in Cantonese, and Kawahara (2002) and Kubozono (1999) for Japanese). Talkers can only produce these connected speech forms if at some point individual segments are available in speech planning (Meyer & Belke, 2007).

A second kind of evidence comes from speech error research. The vast majority of speech sound errors (typically 70%–90%) involve single consonant or vowel phonemes (Dell, 1995; Nooteboom, 1969; Shattuck-Hufnagel, 1983). The high rate of single segment errors can be accounted for if phonological encoding specifically selects segments, because individual segments can occasionally be mis-selected. This assumption also accounts for the rarity of errors involving groups of segments, because they must occur as combinations of independently occurring single segment errors. Encoding words one segment at a time also reveals an interesting dynamic, where a supplanted segment can re-emerge immediately downstream, as in the segmental exchange *h**e**ft-l**e**misphere* for *left-hemisphere* (Fromkin, 1973). Segmental exchanges are naturally explained as reciprocal segmental mis-selections, but they are almost impossible to account for in other ways, such as feature movements or mis-selected words (Dell, 2014).

Phonological features and syllables, on the other hand, present a kind of paradox in the analysis of speech errors (Dell, 1986). They are relevant to the frequency distributions of segmental errors, but they do not seem to be mis-selected. For example, sound errors in which the intruder and intended sounds share all but one feature, and are therefore phonologically similar, are very common (MacKay, 1970; Shattuck-Hufnagel & Klatt, 1979). However, speech errors that suggest the mis-selection or movement

of a feature, as in the anticipation of [labial] in *thumb* → *fumb* (Stemberger, 1983), are relatively rare (Fromkin, 1971) and alternative analyses (e.g., the chance occurrence of mis-selected segments) have not been ruled out. Likewise, segmental errors are sensitive to syllable structure in the sense that intruder sounds tend to retain the syllabic roles that they have in source positions (Boomer & Laver, 1968; Fromkin, 1971). But whole syllable errors are vanishingly rare, at least in Indo-European languages (Bock, 1991; Dell, 1995). For the consensus model sketched above, the influences from syllables and features can be captured with the activation dynamics linking segments to these ingredients and structures, while maintaining the strong assumption that word-form retrieval is a matter of activating and selecting segments, not features or syllables.

The idea that phonological encoding proceeds via the segment also has extensive support from experimental studies. In the picture-word paradigm, for example, the response times for target picture names that are accompanied by phonologically related auditory or printed distractors are reduced relative to targets with unrelated distractors (e.g., Damian & Martin, 1999; Meyer & Schriefers, 1991; Starreveld & La Heij, 1996). Importantly, the phonologically related words in these experiments shared a minimum of two initial segments or two final segments with the target word (e.g., distractor: *muscle*, target: *mushroom*), but the shared elements did not form a coincident syllable. Likewise, masked priming studies have documented priming effects from shared segments independent of whether the shared segments comprise a syllable in the target word (e.g., Schiller, 1998, 2000). The reduced latencies in these priming studies are consistent with the assumption that shared segments of the prime or distractor words support the phonological production of the target word by boosting the activation of some of its component segments.

Further experimental support for segmental encoding comes from form preparation (originally called implicit priming) experiments. In these studies, participants are asked to produce target words in response to pictures, learned cue words (as in cue: *cat* → target: *dog*), or the printed words themselves. The words are tested in blocks in which the set of target words is either phonologically related or unrelated. With Dutch native speakers, Meyer (1990, 1991) found that participants produced the target words more quickly when they shared one or more initial segments, and that the magnitude of the effect on response time was a function of the number of shared segments. Furthermore, simply sharing many features in the set of target word onsets (i.e., phonologically similar but not identical segments) is not enough to allow full form preparation (Roelofs, 1999), so the form preparation effect appears to be driven largely by the number of shared segments.

While the evidence for the fundamental role of segments in phonological encoding is compelling, it should be noted that the evidence against the selection of other phonological structures is not so clear cut. For example, the issue of the activation of syllable level units in masked priming is unresolved (Ferrand et al., 1996), with two studies arriving at different conclusions about the same language (Ferrand et al., 1997; Schiller, 2000). Furthermore, syllable body (CV) skeletal structure, the prosodic structure that abstracts over consonant and vowel phonemes, appears to be relevant to speech error patterns (Stemberger, 1990; Vousden et al., 2000), although the same structure does not seem to be accessible in form preparation experiments (Roelofs & Meyer, 1998). Finally, casual speech phonology involving what appear to be deletion or insertion of segments also has alternative analyses based on the dynamics of gestural coordination, which suggests that at least some of these processes may not, at a deeper level, involve segments (Browman & Goldstein, 1990).

3 | PROXIMATE UNITS

Against this backdrop, a series of studies on Mandarin Chinese led to discoveries about phonological encoding that are strikingly different from Indo-European language findings, and consequently required revision of the consensus model. First, in contrast to languages like English, Chen (2000) found that syllable-sized units can be mis-selected in Mandarin, including mis-selections of atonal syllables (i.e., a whole syllable representation stripped of tone), as in /qing1 zhuo2 du4/ → *qing1 du2 du4* 'clarity (清濁度)'. Because they were collected from recorded conversations, the chance rate of syllable errors could be calculated from the baseline error frequencies. Chen found that even with rather conservative assumptions, syllable errors occur at rates well above chance in Mandarin. Recent work investigated syllable errors using similar methods in a related Chinese language, Cantonese, and again found a significant presence of such errors (Alderete, 2022).

Building on the fact that whole syllables can slip, Chen et al. (2002) investigated form preparation in Mandarin with the hypothesis that syllables may have a more salient role in phonological encoding than in other languages. In particular, they constructed sets of disyllabic response words with shared word-initial material, contrasting shared initial syllables (either tonal or atonal) with shared initial onset consonants. They found that both shared syllables and atonal syllables showed a form preparation benefit, measured in naming time, though the effect of shared atonal syllables was naturally smaller. Shared onsets, by contrast, did not show any benefit, another clear contrast with English and Dutch, where shared initial onset consonants show a preparation benefit (Meyer, 1991; Roelofs, 1999). This difference between English and Mandarin languages was replicated and extended to monosyllabic conditions in O'Séaghdha et al. (2010). Subsequently, the key contrast between shared initial syllables and onsets was also investigated in Cantonese, which was found to pattern like Mandarin (Wong & Chen, 2008, 2009, but see Wong & Chen, 2015 also for apparent cases in which combinations of segments with tone can also have priming benefits). Clearly, syllables are processed in Chinese languages rather differently than they are in Indo-European languages.

To account for the differences between English and Dutch, on the one hand, and Chinese languages, on the other, O'Séaghdha et al. (2010) advanced the Proximate Units Principle. The idea is that the initial step of word-form encoding differs cross-linguistically in the type of the units first retrieved from memory (see Figure 1). The proximate unit is the first selectable unit at the phonological stage of encoding. In Mandarin, the first selectable unit after lexical selection is the syllable, which in turn leads to the activation of segments downstream. Because syllables are actively selected, they may be mis-selected, accounting for syllable errors. Likewise, because they are activated first in form encoding in Mandarin, they constrain what a speaker can plan in advance and can conceive as a starting point for a word (O'Séaghdha & Chen, 2009; O'Séaghdha et al., 2010). The reason response words that only share the same onset (e.g., ***m****o1-yu2*, ***m****a2-zui4*, ***m****u3-dan1*) are not typically planned in advance in Mandarin is because these are viewed by Mandarin speakers as different syllables, so the shared [m] is not deployed as a word onset for a set of response words. In English, on the other hand, the segment is the proximate unit, so syllable errors are not predicted as mis-selections of syllable level chunks and form preparation is mediated by segments.

While speech error and form preparation evidence were critical to initial support for proximate units, the principle was also supported by convergent data from priming experiments. Given the fundamental role of syllables in Mandarin and Cantonese, researchers asked if atonal syllables produce priming effects in the masked priming and picture-word paradigms discussed above. In masked priming studies of Mandarin, response latencies were shorter when the initial syllable matched the prime atonal syllable, for example, /fa3/ → *fa1-nan4*, relative to the same segmental string but with a mis-match in syllables, as in, /fa3/ → *fan1-an4* (Chen et al., 2003, 2016; Lin & Chen, 2003). Crucially,

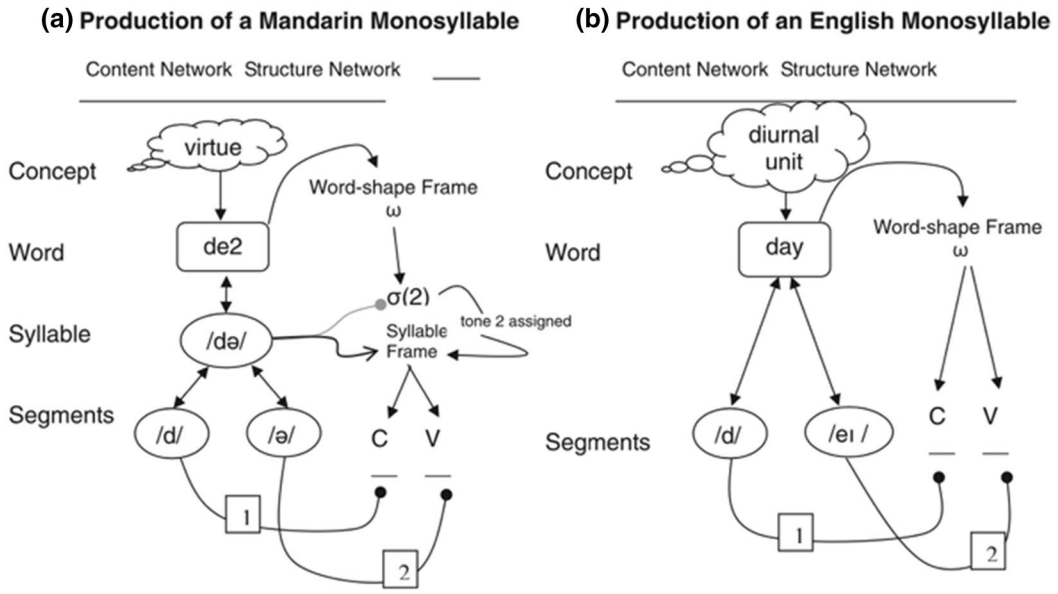

FIGURE 1 Proximate units for simple CV monosyllables in (a) Mandarin and (b) English. *Source*: From O'Séaghdha et al. (2010)

Chen et al. (2016) showed, using pictures rather than characters to elicit target productions, that onsets alone do not prime in Mandarin, but syllables do. These findings are consistent with those of Wong and Chen (2008) for Cantonese, discussed above, showing that syllables, but not onsets, have facilitatory priming effects in the picture-word paradigm. If form encoding in these languages initially involves activating the first syllable of a word, both priming effects can be accounted for as the pre-activation of the first syllable of the target word.

The view that phonological encoding may begin with phonological structures other than the segment is also supported by evidence from Japanese, which differs in important ways from both Chinese and Indo-European languages. It has long been argued that the mora is a fundamental unit of processing in Japanese (Otake et al., 1993; Tamaoka & Makioka, 2009). For this discussion, we conceive of morae as being smaller than syllables but sometimes having more than one segment. Thus, in addition to coda consonants, the first CV sequence of a syllable or word is a mora as well (Kubozono, 1985; McCawley, 1978). While such errors are rare in English and German, sound errors consisting of one or two phonemes are more common in Japanese, as in /no-ren ni u-de-o-si/ → ***u****-de.n ni* ***no****-re-o-si* 'waste of labour', where two strings dominated by a single mora, *u* and *no*, exchange. While this error is akin to a syllable exchange, others involve a CV exchange with a nasal coda, as in /da.teN saN no ha.ra/ → *da.teN sa.****ra*** *no ha.ra* 'Hara with three hitting points' (Terao, 2022), which cannot be a syllable error. Such cases have been attributed to the mora in Japanese (Kubozono, 1989; Nakayama & Saito, 2014; Saito & Inoue, 2017), suggesting that it is a selectable unit in this language. Furthermore, parallelling the differential pattern found by Chen et al. (2002) and O'Séaghdha et al. (2010) with Mandarin syllables, form preparation experiments with Japanese participants found that initial shared onsets do not reduce latencies but initial CV sequences do (Kureta et al., 2006). Together these two findings support an extension of the Proximate Units Principle to include the mora as a proximate unit in Japanese and thus suggest that it may have broad application across diverse languages.

Research has recently expanded to other East Asian and Southeast Asian languages, including Korean (Han & Verdonschot, 2019; Li et al., 2022; Verdonschot et al., 2021), Vietnamese

(Verdonschot et al., 2022), and Thai (Winskel & Ratitamkul, 2020). This research is on-going, but results to-date suggest parallels in terms of the proximate units discussed above. Verdonschot and colleagues found syllable, but not syllable onset, priming effects in picture-word distractor experiments in Korean, suggesting that Korean patterns with Chinese languages and Japanese, while having some distinctive attributes (e.g., word-internal resyllabification). Li et al. (2022) reported similar results for Korean form preparation, but suggested the CV rather than the whole CVC unit may be the critical preparation unit, which is consistent with the role of phonotactics within the syllable body (Lee & Goldrick, 2008). Thus, these findings could comport with mora rather than syllable proximate units. In contrast, Verdonschot et al. (2022) argue that Vietnamese deploys proximate phonemic segments, like Indo-European languages. Likewise, Winskel and Ratitamkul (2020) concluded, based on masked priming evidence, that phonemes are primary units in Thai.

With regard to the different conclusions about Korean (i.e., the syllable vs. CV and coda units), it is interesting to consider the conclusion of Wang et al. (2018) based on picture-word distractor priming evidence that both syllable and CV units are employed in Cantonese, but that the former are retrieved first. The debate concerning the possibility of multiple units in Korean and Cantonese suggests that initial phonological units are not necessarily monolithic within every language. In general, the Proximate Units Principle concerns the nature of memory for lexical forms. It suggests that phonological chunks that are sufficiently consistent during encoding (as in segments, syllables, or the variable shapes of morae CV/V/C) may become the predominant proximate units of a language. Both morae and syllables are variable in nature, so their chunking in memory must also be variable. Diphthongs provide another familiar example of variable chunking where complex vowels may plausibly be retrieved wholistically as simple vowels are. While we are not aware of any relevant evidence, we may speculate that frequent or exclusive co-occurrence of consonants in particular clusters could also lead to emergence of corresponding proximate units in phoneme-centric languages. For more discussion on what may motivate and constrain variation in proximate units, both cross-linguistically and language internally, see Chen et al. (2002) and O'Séaghdha (2015).

4 | SERIAL ORDER AND STARTING POINTS

The serial order problem, or how the elements of complex plans are sequenced, is a fundamental problem in phonological encoding, and indeed in the larger field of cognitive science (Dell et al., 1997; Lashley, 1951; Shattuck-Hufnagel, 1979). Simple one-word productions, such as *cat*, involve not only selecting the correct sounds, but also creating a speech plan in which these sounds occur in the correct order. Recognising hierarchical structure in speech planning is a common way of addressing this problem (see e.g., Dell, 1986; Sevald et al., 1995; Smolensky et al., 2014). For example, if the component sounds of *cat* are assigned syllabic roles, and then inserted into a syllable template, onset [k] will appear before coda [t], because onsets appear before codas in the template. However, this structure is not enough to order the segments of polysyllabic words or larger phrases. Complex plans such as these can be broken down into sub-plans and aligned dynamically with the talker's attention by using a so-called primacy gradient (Dell et al., 1997; Nozari & Dell, 2012; Pinet et al., 2019). In producing the phrase *cat walk*, for example, both words are activated together, but *cat* receives higher activation while the talker is attending to this word. As a result, onset [k] of *cat* will have higher activation than a coda [k], which is also activated by *walk* but initially has a weaker activation. When attention then shifts to the next word, the activation of *cat* is decreased and *walk* is increased, setting up the correct selection of its component sounds in their proper positions, including coda [k]. We explore here how

these ideas, hierarchical structure and primacy gradients, contribute to serial order effects in diverse languages.

An important aspect of the serial order problem in phonological encoding is the treatment of phonological material at the beginning of words. Multiple lines of evidence suggest that word onsets are processed differently from non-initial phonological material. In speech error studies, for example, word onsets differ from the rest of the word in that they are more error-prone in both spontaneous and experimental speech errors (see Wilshire (1998) for review). Interestingly, and perhaps paradoxically, evidence also exists that word onsets are more robust in some production behaviours. For example, in tip-of-the-tongue states (i.e., states in which the talker has selected a lemma, but cannot retrieve its form), talkers nonetheless are more successful at retrieving word onsets than other portions of the word (Brown, 1991; Brown & McNeill, 1966). As discussed above, word onsets are also pivotal in form preparation. Only shared material at the beginning of a word makes a preparation benefit possible (Meyer, 1990, 1991), and the shared material must include the word onset. Finally, word onsets are more likely to be recovered than non-initial material by aphasic patients with impairments affecting phonological processing (Kohn & Smith, 1994; Wilshire & McCarthy, 1996). Clearly, word beginnings must be specifically addressed in any satisfying account of serial order in phonological encoding, though importantly, what constitutes a word beginning may differ from language to language.

A variety of linguistic and psycholinguistic accounts have been suggested for the special status of word onsets, but one account is especially compelling as a general solution to the problem. This solution involves a kind of within-word primacy gradient using a control signal for the starting points and ending points of words (Houghton, 1990; Jacobs & Dell, 2014; O'Séaghdha, 2015). Houghton's original competitive cueing model posited Start and End nodes that mediate between word and phoneme representations, and a system in which the activation values sent to phonemes from these nodes change over time. In particular, the Start node initially sends higher activations to segments in early positions than to later segments (depicted by line strength in Figure 2), but this preference decreases as processing moves on to later segments. As the activation signal from the Start node decreases and levels out, activation from the End node increases, supporting selection of final sounds.

Houghton (1990) specifically draws on the special status of word onsets to motivate the begin-to-end bias in activation achieved with the Start nodes. In particular, the activation histories generated by his model show that word onset segments have greater activations than later phonemes earlier in word production, which is consistent with the evidence that word onsets are more readily

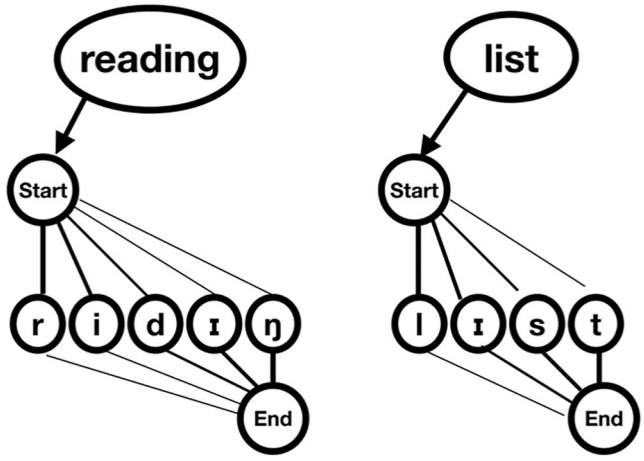

FIGURE 2 Phonological encoding with starting points

recoverable in tip-of-the-tongue states and in disordered speech. Though this has not been tested computationally, Houghton also suggests that the use of Start nodes can help with the analysis of the word onset effect in speech errors. In corpus studies, segmental errors in initial positions occur well above chance levels, constituting between 50% and 90% of all single segment errors (MacKay, 1970; Shattuck-Hufnagel, 1987). The word onset effect has also been found in experiments designed to distinguish between initial and non-initial elements (Shattuck-Hufnagel, 1992; Wilshire, 1998). An important finding relevant to starting points is that the word onset effect is limited to contextual errors; the effect goes away in errors in which intruder sounds do not come from the linguistic context (Wilshire, 1998). Wilshire uses this constraint to motivate two requirements of any analysis of the effect: (a) competition in activation between words that resolves itself over time, and (b) continuous communication between lexical and phonological levels. The reasoning is that the word onset effect depends on competition between words that cascades down to the phonemic level, and this competition is only salient in multiword contexts.

A model of phonological encoding with starting points meets these criteria, and in doing so, it suggests a viable account of the word onset effect. Returning to Figure 2, the phrase *reading list* activates both lemmas, *reading* and *list*, with the former receiving more activation initially. The jolts of activation cascade down to the Start nodes from both of these words, and then to the component segments. Importantly, more activation accrues to the onsets because of the start-to-end bias inherent to these links. Thus, funnelling more activation to onsets creates a competition between onset [r] and the highly similar onset [l] in multi-word phrases, which can lead to mis-selections and errors such as ***l****eading list* and *reading* ***wr****ist*. There will also be competition in non-initial segments, which can lead to errors with sounds in non-initial locations. But competition will be highest between initial segments because of the role of the Start nodes, thus accounting for both the specific position affected in the word onset effect, together with its contextual nature.

The word onset effect in speech errors is cross-linguistically robust. Above chance rates of errors in word initial position have been found in German (Berg, 1991), two large studies of English (Shattuck-Hufnagel, 1987; Vousden et al., 2000), Cantonese (Alderete, 2022), and Korean (Han et al., 2019). However, Spanish sound errors appear to contravene this consensus. Berg (1991), investigating the del Viso error corpus (del Viso et al., 1991), showed that contextual word-initial errors had a 27.3% occurrence, well below the chance rate of approximately 40%, and that corresponding word-medial errors occurred at above chance rates. This under-representation of word onset errors, and over-abundance of non-initial errors, was also found in another large study of Spanish (Pérez et al., 2007), though not specifically in contextual errors, which are crucial to establishing the lack of word onset effect (Wilshire, 1998).

We suggest that further study of Spanish is required to understand these patterns. We note that Spanish phonological processing is unusual in several ways that may relate to the absence of a clear word onset effect. Investigation of syllable frequency in the del Viso corpus also revealed that, contrary to expectations, error, target and also source syllable frequencies were significantly lower than chance levels (Santiago et al., 2007). In addition, Spanish words from sparse lexical networks were spoken faster than comparable words from dense lexical networks, the exact opposite pattern found in English (Vitevitch & Stamer, 2006). Pursuing this further, Vitevitch and Stamer probed the densities of four-segment words in the two languages and found that the English words had more neighbours based on the first two segments, whereas Spanish had the opposite pattern, with more neighbours sharing the last two segments. In other words, it seems that the beginnings of words are more confusable in English than Spanish (see Griffin and Crew (2012) for similar rationale). We may thus conjecture that the apparent protection of word onsets may be due to their lower confusability relative to the rest of the word than in other Indo-European languages.

Returning to the role of starting points, models of phonological encoding that include them provide the basis for explanation in a variety of form preparation effects (Jacobs & Dell, 2014; O'Séaghdha, 2015). The basis of form preparation is that talkers can activate representations of initial portions of the related (homogeneous condition) target words that start with the same initial segment or sequence (Levelt et al., 1999). While this accounts for English mono-morphemic word preparation, it is not sufficient to account for null findings using the same research paradigm for dis-assembled English compounds. Jacobs and Dell (2014) showed that when compounds like *sawdust* are broken apart, as in, cue: *saw* → target: *dust*, responses latencies for targets with shared material (e.g., *out* → [d]*oor, touch* → [d]*own*) were not reduced relative to the baseline heterogeneous condition. On the basis of this finding, Jacobs and Dell argued that form preparation requires that the shared phonological representations begin with a starting point, because this gives a consistent slot for talkers to anchor the phonological representation. There is but one starting point in a compound (because it is one phonological word), so a [d] that appears medially, as in *saw*[d]*ust*, cannot be planned in advance.

These assumptions can be combined with cross-linguistic differences in proximate units to account for the striking differences between English and Mandarin (O'Séaghdha, 2015). As discussed above, while both English and Mandarin exhibit form preparation effects for initial (atonal) syllables, Mandarin differs from English in that initial segments and CV sequences do not exhibit the effect in Mandarin. The difference can now be described in terms of proximate units. Segments are the proximate unit for English, and so by assuming starting points are segments, we account for the fact that shared initial segments can produce the priming effect. However, syllables are the proximate unit in Mandarin, so the starting points for Mandarin words are syllables, accounting for the fact that shared initial syllables, but not shared initial segments, allow form preparation. Interestingly, experiments with disyllabic compounds in Mandarin provide additional evidence for the idea that Mandarin words are sequenced in terms of syllables rather than segments (Chen & Chen, 2015; O'Séaghdha et al., 2010). When disyllabic compounds are broken up into cue-targets, as in *lian2* → *jie1* (from *lian2jie1* 'connect') and combined in the homogenous set with other compounds sharing the same segmental material in the second syllable (e.g., *chai1* → *jie3*), form preparation occurred. The important insight, made explicitly in Jacobs and Dell (2014), is that while the compounds did not share an initial syllable, Mandarin speakers know how to build a representation because words are ordered by syllable chunks, so they can reliably place an atonal syllable (*jie* in this case) in the second slot. This planning problem contrasts with the case of non-initial segments in English, where medial segments do not consistently appear in a predictable slot (e.g., [d] in the third slot for *sawdust* [sɑd̲ʌst], but fourth for *touchdown* [tʌtʃd̲aʊn]), and so they cannot be planned in advance.

5 | THE FUNCTION AND TIMING OF SYLLABLE STRUCTURE

Syllable structure has different functions in different models, and it is also introduced at different points in time in phonological planning. Virtually all models implement an online process of syllabification that assigns selected segments to syllabic roles, as in Dell's (1986) well-known use of a syllable frame to syllabify segments (see also Sevald et al., 1995; Shattuck-Hufnagel, 1979; Smolensky et al., 2014). Syllabification in this case both orders segments within a syllable and provides the basis for (learned) constraints on phonological encoding, such as phonotactic constraints (Alderete & Tupper, 2018; Dell et al., 2000, 2021; Goldrick & Daland, 2009; Vousden et al., 2000). In some models, syllables are represented as chunks and play a role of mediating the activation flow between a word and its component sounds. We have already seen syllables function this way in languages where the syllable is the proximate unit (e.g, Mandarin, Section 3), and indeed syllables may have

this mediating role in other languages where syllable structures are assumed to be part of the lexical network, but not actively selected (see e.g., Dell, 1986; Romani et al., 2011; Santiago et al., 2007; cf., Sevald et al., 1995). In addition, syllables are also represented in abstract prosodic frames to encode metrical information, and they have been hypothesised to package articulatory planning of whole syllables in so-called syllable programme nodes (Levelt et al., 1999), though these articulatory programme nodes are different in kind than syllables at the phonological encoding level.

As alluded to in Section 2, the reference to syllable structure and whether it is lexicalised is a long-standing and unresolved issue. One perspective is that, while segments are organised into syllables and referred to in prosodic frames, they are not lexicalised in the sense that syllabic information is tied to individual words or morphemes (for review, see Meyer & Belke, 2007). In support of this view, its advocates argue that lexicalised syllabification would need to be re-syllabified in connected speech, as well as to the general absence of evidence for syllable structure in priming and form preparation studies (Roelofs & Meyer, 1998; Schiller, 1997). On the other side of this debate, others have argued, based on speech error evidence, phonological deficits, and priming evidence, that syllables are linked and co-activated with lexical items in the lexical network representing knowledge of words (see Romani et al., 2011 for review). Finally, more nuanced models entertain multiple levels of production processing, allowing reference to syllables at later stages of phonological processing, though without lexicalised syllable chunks per se (Goldrick & Rapp, 2007; O'Séaghdha & Marin, 2000).

One context for probing the lexicalised status of syllables is re-syllabification, the propensity for segments to assume different syllabic roles in citation and connected speech forms. Because English and Dutch exhibit re-syllabification in a variety of speech contexts (Schiller et al., 1996), Levelt (1992; Levelt et al., 1999) argued that rather than being stored in lexical representations, syllabification is an online generative process. If syllable structure is lexically stored, a large number of word-final consonants will require re-syllabification when they are followed by vowel-initial words. For example, in *neglect it*, the [t] is a coda in the citation form but a syllable onset in connected speech: [nə.glɛk.tɪt]. Such re-syllabifications are inefficient and unnecessary if syllable structure can be computed online together with other metrical analyses needed for stress patterns.

How do the facts of languages beyond Indo-European contribute to this debate? When we broaden the empirical scope, the argument against lexicalised syllables is not as forceful in languages that appear to disfavour re-syllabification. For example, Italian words are predominantly vowel-final, and so lack final consonants that can re-syllabify with vowel-initial words (Romani et al., 2011). Likewise, many have noted that Chinese languages like Mandarin and Cantonese, as well as Hindi, lack re-syllabification of segments in content words (Chen, 2000; Myers, 2010; Ramoo et al., 2021). In these languages, it would seem that reservations based on computational efficiency are not relevant, making lexical specification of syllables more plausible. Indeed, the apparent lack of re-syllabification is part of the motivation for the privileged status of encoding syllables in Chinese languages (Chen, 2000).

It turns out, however, that even languages that disfavour re-syllabification of codas admit it in other environments. For example, while Italian lacks re-syllabification of word-final coda consonants, it has a well-defined process of phrasal phonology that is a kind of 're-syllabification in reverse'. *Raddopiamento sintactico* is a process that geminates a word-initial onset before stressed word-final vowels (e.g., /ki#soːno/ → [kis.soːno] *chi sono* 'Who are they?'), which is standardly assumed to involve re-syllabifying an onset consonant as an ambisyllabic consonant linked to both the final mora of the first syllable and onset position of the second (Repetti, 1991). Likewise, while Mandarin Chinese lacks re-syllabification in many of the contexts where it would be predicted in English or Dutch, it does nonetheless have gemination of coda consonants before certain clitics, as in /næn#a/ → [næn.na] 'Hard!', a kind of progressive *raddopiamento sintactico* which again involves creating an ambisyllabic consonant that straddles both coda and onset positions (Duanmu, 2007; Lin, 2007). In this

context, it is interesting to note that, while re-syllabification is pervasive in English connected speech, a significant portion of cases seems to be associated with specific function morphemes such as *and*, *it*, and *of*. For example, Ramoo et al. (2021) investigated re-syllabification in a large English corpus and found that 30% of the cases involved just two phrases, *and uh* and *and it*. Thus, languages like Mandarin and Italian seem to lack re-syllabification, but nonetheless have re-assignment of syllabic roles in specific contexts. Conversely, the case of re-syllabification in English, though it differs from Mandarin and Italian by re-parsing segments in content words (e.g., *neglect it as* [nə.glɛk.tɪt]), also seems to be more predictable than previously thought in the sense that it is significantly associated with the phrasing of functional items, which compares to the situation in Mandarin and Hindi. Given that none of the languages reviewed above absolutely prohibit re-syllabification, nor allow it to apply in an unrestricted fashion, we think that future work can benefit from more attention to the specific prosodic and morpho-syntactic environments that support re-syllabification.

The structure of speech errors is also relevant to the question of when syllable structure enters into phonological planning. In particular, the lack of sensitivity to syllable structure in Arabic, relative to Germanic languages like English and German, has been argued to require language-particular differences in the emergence of syllable structure in phonological encoding. Arabic differs from Germanic languages in having a far higher rate of within-word errors (80.5% Arabic, cf. 13.7% English and 6.4% German) relative to other between-word contextual errors (Berg & Abd-El-Jawad, 1996). In addition, these within-word errors in Arabic are not subject to psycholinguistic effects commonly observed in Indo-European languages, namely the syllable context constraint (i.e., intruder and source sounds have the same syllabic role) and the onset bias, but they are obeyed in between-word errors. This contrasts with English and German, where both within- and between-word sound errors obey these constraints. Berg and Abd-El-Jawad connect these facts to the non-concatenative nature of Arabic morphology, which, because consonants and vowels need to be intercalated to form words, prevents early syllabification of segments. For example, the first consonant in the triliteral root *k-t-d* 'to write' is syllabified as an onset in *kitaabun* 'book', but a coda in *maktab* 'office', and so determining a specific syllabic role for individual consonants requires first associating the root with the appropriate CV pattern and vowels. Phonological encoding in Arabic therefore has a longer time interval in which segments are selected, but have not yet been syllabified, which accounts for the lack of sensitivity to this structure in 'early' within-word errors. Consonants and vowels do not need to be intercalated in German and English, and so syllabification is more immediate in these languages, as shown by their sensitivity to syllable structure in within-word errors. We have seen that general theories of phonological encoding differ in when syllable structure emerges in phonological planning (e.g., Levelt et al., 1999, cf. Romani et al., 2011, Shattuck-Hufnagel, 1979 et seq.), but what this comparison shows is that emergence of syllable structure itself has language particular variation.

A related question is whether syllable structure emerges from online processing, or is specified in the mental lexicon in the sense that individual lexical items are endowed with syllable structure, as argued explicitly by Romani et al. (2011). Speech errors from Mandarin, Japanese, and Cantonese provide relevant evidence. First, glides and high vowels in Mandarin interact with each other in ways that suggest they are unspecified for syllabic roles in early production processing and accommodate to new positions (Wan, 1997). These are in effect violations of the consonant-vowel constraint, whereby consonants interact with consonants, vowels with vowels (Shattuck-Hufnagel, 1986). For example, in the perseveration of the pre-nuclear glide [ʊ] to a nucleus slot in /ʊ̯ɛn21 li35/ → ʊ̯ɛn21 l[**y**]35 'far away', the glide becomes the related front round vowel [y]. If glides and vowels were pre-compiled for syllable roles, we would expect some resistance to these accommodations, but Wan gives 31 examples of this interaction and it is exceptionless. Japanese and Cantonese have similar violations of the consonant-vowel constraint, except they occur in the coda position (for Japanese see Kubozono, 1989,

Cantonese, Alderete, 2022). Coda consonants can be re-assigned to the second position of a vowel nucleus, as in /tsau22/ → tsa<u>k</u>22 就 'then', and vice versa (vowels swap with consonants), which again is unexpected if these segments are locked into pre-determined syllabic roles (see Kubozono (1989) on the role of the mora for establishing the slot for this substitution). Thus, while there may be some compelling evidence for lexically specified syllable structure, a cross-linguistic account of speech error patterns presents interesting problems for it.

6 | FUTURE DIRECTIONS

Our goal was to re-evaluate basic assumptions about phonological encoding in light of evidence from a broader cross-linguistic sample. Our findings can be summarised as follows:

1. Languages have different proximate units, defined as the first selectable phonological units activated after lexical selection
2. Phonological encoding that implements a primacy gradient with starting points can account for the special status of word onsets, including the word onset effect and the initialness requirement on form preparation
3. Proximate units may be designated as starting points, accounting for differences in the linguistic level of word onset effects
4. While the structure of some languages works against inter-word re-syllabification, all languages studied exhibit re-syllabification in at least some linguistic contexts
5. Speech error evidence in Arabic, Japanese, and Chinese languages supports online processes of syllabification

While we have attempted to summarise research on languages that have not figured prominently in dominant models of phonological encoding, we acknowledge that this is just the tip of the iceberg of the evidence that can be uncovered, and new insights can come from exploring other lesser-studied languages with different linguistic and typological make-ups.

How might these findings be pursued in future research? One of the major findings is the importance of the activation and selection of syllables in phonological planning in languages like Mandarin and Cantonese. Given that past research has argued for the prominence of segments based on their involvement in casual speech phonology (Meyer & Belke, 2007), we wonder if similar evidence might be adduced for a role for syllables. Recent research on processes of syllable fusion in Chinese languages is relevant to this question (Kuo, 2010; Lee, 2003; Wong, 2006). In Cantonese, for example, the contents of two syllables can be merged into a single syllable, as in /dzi55 dou33/ → *dziu53* 'know' (Wong, 2006), in this case retaining the initial of the first syllable and part of the second rime. Unlike contraction rules and word blends, syllable fusion is a rate dependent process affected by morpho-syntactic features and word frequency, and it can result in emergent phonological structures, such as diphthongs, that are otherwise unattested in the language. Operationally, it appears that syllable fusions involve deletion of unstressed syllables and the alignment of the resulting segments to a prosodic template. Thus, just as casual speech phonology at the segmental level motivates segments as primary units in Indo-European languages, casual speech at the syllabic level motivates syllables as primary encoding units in Chinese languages. We think it will be worthwhile attempting to relate the existence of such syllable level phonetic processes with the syllable as a proximate unit, which seems to fit with our current knowledge of syllable fusion in Chinese languages.

Another future direction is to consider the impact of the prominence of syllables on phonological neighbourhood structure. Network science has provided a number of measures for gaging the interconnectedness of lexical structures that have proved useful in accounting for a range of language phenomena (see e.g., Vitevitch & Castro, 2015 for review). The basic network structure developed for Indo-European languages involves a notion of word-level phonological similarity based on the segment. That is, lexical items are linked if they are phonological neighbours in the sense that the two words differ in a change (i.e., deletion, addition, or substitution) of a single segment. For example, density in these segment-based similarity networks (the number of neighbours a word has in the network) has been shown to have an inhibitory effect in word comprehension (Luce & Pisoni, 1998; Vitevitch & Luce, 1998) and a facilitative effect in word production (Chen & Mirman, 2012; Gordon & Dell, 2001; Vitevitch, 1997; Vitevitch, 2002; cf. Sadat et al., 2014).

Our review above has highlighted the importance of syllables in production processing, leading to the question of whether network structure might be defined in terms of shared syllables rather than shared segments (see O'Séaghdha, 2015). For example, how would a lexical network be defined in terms of syllable units? Will it require at least one shared syllable or all but one, as with segments? Are syllabic neighbours defined with or without tone? Turning to similarity among syllables themselves, preliminary work using a minimal pair word production task in Mandarin has considered a range of possible networks based on form similarity of entire syllables, syllable subconstituents, and segment-based similarity, and found network effects based on a combination of segment and tonal similarity, but not syllable-based similarity (Neergaard & Huang, 2019). Future research can draw on the database for this project (Neergaard et al., 2016) to investigate new behaviours and also consider the role of multiplex representations that combine different dimensions of form similarity, as has been successfully implemented for form and meaning similarity (Castro et al., 2020).

Another emergent question concerns linking up the two theoretical threads for proximate units and starting points. In Mandarin, the proximate unit is a syllable, and the syllable is the starting point for words, as confirmed in speech errors and experimental tests. But the selection of segments still proceeds, syllable by syllable, as well (see Figure 1), and indeed speech error evidence in Chinese languages requires syllable-internal segmental encoding in addition to syllable encoding (Alderete et al., 2019). Does syllable-internal encoding also involve a primacy gradient, with Start nodes enhancing activation of syllable-initial segments? If so, when does this activation commence and what is its scope in planning other syllables within and across word boundaries? As in other languages, privileging of starting points within syllables could account for onset movement errors and the onset bias. We know of no work that has looked for this bias in non-initial syllable onsets in Chinese languages, related to Berg's (1991) finding of such effects in Spanish. We therefore think it will be fruitful to investigate how the dynamics of both word-level and syllable-level activation gradients combine in Mandarin and Cantonese, and if they indeed explain the distribution of segmental and other errors in these languages.

ACKNOWLEDGEMENTS

We gratefully acknowledge Shigeto Kawahara and Shin-ichi Tanaka for helpful comments on earlier versions of this work. Our research was supported in part by a grant from the Social Sciences and Humanities Research Council of Canada (435-2020-0193).

ORCID

John Alderete https://orcid.org/0000-0002-6741-9426

REFERENCES

Alderete, J. (2022). Cross-linguistic trends in speech errors: An analysis of sub-lexical errors in Cantonese. *Language and Speech*, *2022*, 002383092110710. https://doi.org/10.1177/00238309211071045

Alderete, J., Chan, Q., & Yeung, H. H. (2019). Tone slips in Cantonese: Evidence for early phonological encoding. *Cognition*, *191*, 103952. https://doi.org/10.1016/j.cognition.2019.04.021

Alderete, J., & Tupper, P. (2018). Phonological regularity, perceptual biases, and the role of grammar in speech error analysis. *WIREs Cognitive Science*, *9*, e1466. https://doi.org/10.1002/wcs.1466

Anand, P., Chung, S., & Wagers, M. (2011). Widening the net: Challenges for gathering linguistic data in the digital age. NSF SBE 2020: Future research in the social, behavioral, & economic sciences.

Bauer, R. S., & Benedict, P. K. (1997). *Modern Cantonese phonology*. Mouton de Gruyter.

Berg, T. (1991). Phonological processing in a syllable-timed language with pre-final stress: Evidence from Spanish speech error data. *Language and Cognitive Processes*, *6*, 265–301. https://doi.org/10.1080/01690969108406945

Berg, T., & Abd-El-Jawad, H. (1996). The unfolding of suprasegmental representations: A cross-linguistic perspective. *Journal of Linguistics*, *32*, 291–324. https://doi.org/10.1017/s0022226700015905

Bock, K. (1991). A sketchbook of production problems. *Journal of Psychological Research*, *14*, 382–391.

Bock, K., & Ferreira, V. S. (2014). Syntactically speaking. In M. Goldrick, V. S. Ferreira, & M. Miozzo (Eds.), *The Oxford handbook of language production* (pp. 21–46). Oxford University Press.

Bock, K., & Levelt, W. J. M. (1994). Language production. Grammatical encoding. In M. A. Gernsbacher (Ed.), *Handbook of psycholinguistics* (pp. 945–984). Academic Press.

Boomer, D. S., & Laver, J. D. M. (1968). Slips of the tongue. *International Journal of Language & Communication Disorders*, *3*, 2–12. https://doi.org/10.3109/13682826809011435

Browman, C., & Goldstein, L. (1990). Tiers in articulatory phonology, with some implications for casual speech. In M. Beckman & J. Kingston (Eds.), *Papers in laboratory phonology I: Between the grammar and physics of speech* (Vol. 1, pp. 341–397). Cambridge University Press.

Browman, C., & Goldstein, L. (1992). Articulatory phonology: An overview. *Phonetica*, *49*, 155–180. https://doi.org/10.1159/000261913

Brown, A. S. (1991). A review of the tip-of-the-tongue experience. *Psychological Bulletin*, *109*, 204–223. https://doi.org/10.1037/0033-2909.109.2.204

Brown, R., & McNeill, D. (1966). The tip of the tongue phenomenon. *Journal of Verbal Learning and Verbal Behavior*, *5*, 325–337. https://doi.org/10.1016/s0022-5371(66)80040-3

Castro, N., Stella, M., & Siew, C. S. Q. (2020). Quantifying the interplay of semantics and phonology during failures of word retrieval by people with aphasia using a multiplex lexical network. *Cognitive Science*, *44*, e12881. https://doi.org/10.1111/cogs.12881

Chen, J.-Y. (2000). Syllable errors from naturalistic slips of the tongue in Mandarin Chinese. *Psychologia*, *43*, 15–26.

Chen, J.-Y., Chen, T.-M., & Dell, G. S. (2002). Word-form encoding in Mandarin Chinese as assessed by the implicit priming task. *Journal of Memory and Language*, *46*, 751–781. https://doi.org/10.1006/jmla.2001.2825

Chen, J.-Y., Lin, W.-C., & Ferrand, L. (2003). Masked priming of the syllable in Mandarin Chinese speech production. *Chinese Journal of Psychology*, *45*, 107–120.

Chen, J.-Y., O'Séaghdha, P. G., & Chen, T.-M. (2016). The primacy of abstract syllables in Chinese word production. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *42*, 825–836. https://doi.org/10.1037/a0039911

Chen, Q., & Mirman, D. (2012). Competition and cooperation among similar representations: Toward a unified account of facilitative and inhibitory effects of lexical neighbors. *Psychological Review*, *119*, 417–430. https://doi.org/10.1037/a0027175

Chen, T.-M., & Chen, J.-Y. (2015). The phonological planning in Mandarin spoken production of mono- and bi-morpheme words. *Japanese Psychological Research*, *57*, 81–89. https://doi.org/10.1111/jpr.12059

Costa, A., Alario, R.-X., & Sebastián-Gallés, N. (2007). Cross-linguistic research on language production. In G. Gaskell (Ed.), *The Oxford handbook of psycholinguistics* (pp. 531–546). Oxford University Press.

Cruttenden, A. (2014). *Gimson's pronunciation of English* (8th ed.). Routledge.

Damian, M. F., & Martin, R. C. (1999). Semantic and phonological codes interact in single word production. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *25*(2), 345–361. https://doi.org/10.1037/0278-7393.25.2.345

Dell, G. S. (1986). A spreading-activation theory of retrieval in sentence production. *Psychological Review*, *93*, 283–321. https://doi.org/10.1037/0033-295x.93.3.283

Dell, G. S. (1995). Speaking and misspeaking. In L. R. Gleitman & M. Liberman (Eds.), *An invitation to cognitive science, Language* (Vol. 1, pp. 183–208). The MIT Press.

Dell, G. S. (2014). Phonemes and production. *Language, Cognition and Neuroscience*, *29*, 30–32. https://doi.org/10.1080/01690965.2013.851795

Dell, G. S., Burger, L. K., & Svec, W. R. (1997). Language production and serial order: A functional analysis and a model. *Psychological Review*, *104*, 123–147. https://doi.org/10.1037/0033-295x.104.1.123

Dell, G. S., Kelley, A. C., Hwang, S., & Bian, Y. (2021). The adaptable speaker: A theory of implicit learning in language production. *Psychological Review*, *128*, 446–487. https://doi.org/10.1037/rev0000275

Dell, G. S., & O'Séaghdha, P. G. (1991). Mediated and convergent lexical priming in language production: A comment on Levelt et al. (1991). *Psychological Review*, *98*(4), 604–614. https://doi.org/10.1037/0033-295x.98.4.604

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. https://doi.org/10.1037/0278-7393.26.6.1355

del Viso, S., Igoa, J. M., & Garcia-Albea, J. E. (1991). On the autonomy of phonological encoding: Evidence from slips of tongue in Spanish. *Journal of Psychological Research*, *20*, 161–185. https://doi.org/10.1007/bf01067213

Duanmu, S. (2007). *The phonology of standard Chinese*. Oxford University Press.

Ferrand, L., Segui, J., & Grainger, J. (1996). Masked priming of word and picture naming: The role of syllable units. *Journal of Memory and Language*, *35*, 708–723. https://doi.org/10.1006/jmla.1996.0037

Ferrand, L., Segui, J., & Humphreys, G. W. (1997). The syllable's role in word naming. *Memory & Cognition*, *35*, 458–470. https://doi.org/10.3758/bf03201122

Fromkin, V. (1971). The non-anomalous nature of anomalous utterances. *Language*, *47*, 27–52. https://doi.org/10.2307/412187

Fromkin, V. (1973). *Speech errors as linguistic evidence*. Mouton.

Garrett, M. (1980). Levels of processing in sentence production. In B. Butterworth (Ed.), *Language production: Vol. I. Speech and talk* (pp. 177–220). Academic Press.

Goldrick, M. (2002). *Patterns of sound, patterns in mind: Phonological regularities in speech production*. (Doctoral dissertation). Johns Hopkins University.

Goldrick, M., & Daland, R. (2009). Linking speech errors and phonological grammars: Insights from Harmonic Grammar networks. *Phonology*, *26*, 147–185. https://doi.org/10.1017/s0952675709001742

Goldrick, M., & Rapp, B. (2007). Lexical and post-lexical phonological representations in spoken production. *Cognition*, *102*, 219–260. https://doi.org/10.1016/j.cognition.2005.12.010

Goldstein, L., Pouplier, M., Chena, L., Saltzman, E. L., & Byrd, D. (2007). Dynamic action units slip in speech production errors. *Cognition*, *103*, 386–412. https://doi.org/10.1016/j.cognition.2006.05.010

Gordon, J. K., & Dell, G. S. (2001). Phonological neighborhood effects: Evidence from aphasia and connectionist modeling. *Brain and Language*, *79*, 21–31.

Griffin, Z. M., & Crew, C. M. (2012). Research in language production. In M. J. Spivey, K. McRae, & M. F. Joanisse (Eds.), *The Cambridge handbook of psycholinguistics* (pp. 409–425). Cambridge University Press.

Han, J.-I., Oh, J., & Kim, J.-Y. (2019). Slips of tongue in the Seoul Korean Corpus of spontaneous speech. *Lingua*, *220*, 31–42. https://doi.org/10.1016/j.lingua.2019.01.001

Han, J.-I., & Verdonschot, R. G. (2019). Spoken-word production in Korean: A non-word masked priming and phonological Stroop task investigation. *Quarterly Journal of Experimental Psychology*, *72*, 901–912. https://doi.org/10.1177/1747021818770989

Harley, T. A. (1984). A critique of top-down independent level models of speech production: Evidence from non-plan-internal speech errors. *Cognitive Science*, *8*, 191–219. https://doi.org/10.1207/s15516709cog0803_1

Houghton, G. (1990). The problem of serial order: A neural network model of sequence learning and recall. In R. Dale, C. Mellish, & M. Zock (Eds.), *Current research in natural language generation* (pp. 287–319). Academic press.

Jacobs, C., & Dell, G. S. (2014). 'Hotdog', not 'hot dog': The phonological planning of compound words. *Language, Cognition and Neuroscience*, *29*, 512–523. https://doi.org/10.1080/23273798.2014.892144

Kawahara, S. (2002). *Similarity among variants: Output-variant correspondence*. (Honors thesis). International Christian University.

Kohn, S. E., & Smith, K. L. (1994). Distinctions between two phonological output deficits. *Applied Psycholinguistics*, *15*, 75–95. https://doi.org/10.1017/s0142716400006986

Kubozono, H. (1985). Speech errors and syllable structure. *Linguistics and Philology*, *6*, 220–243.

Kubozono, H. (1989). The mora and syllable structure in Japanese: Evidence from speech errors. *Language and Speech*, *32*, 249–278. https://doi.org/10.1177/002383098903200304

Kubozono, H. (1999). *Japanese phonetics: An introduction to modern Linguistics Nihongo no Onsei: Gendai Gengogaku Nyuumon 2*. Iwanami.

Kuo, G. (2010). Production and perception of Taiwan Mandarin syllable contraction. *UCLA Working Papers in Phonetics*, *108*, 1–34.

Kureta, Y., Fushimi, T., & Tatsumi, I. F. (2006). The functional unit of phonological encoding: Evidence for moraic representation in native Japanese speakers. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *32*, 1102–1119. https://doi.org/10.1037/0278-7393.32.5.1102

Laganaro, M. (2019). Phonetic encoding in utterance production: A review of open issues from 1989 to 2018. *Language, Cognition and Neuroscience*, *34*, 1193–1201. https://doi.org/10.1080/23273798.2019.1599128

Lashley, K. S. (1951). The problem of serial order in behavior. In L. A. Jeffress (Ed.), *Cerebral mechanisms in behavior* (pp. 112–136). Wiley.

Lee, K. G. (2003). *Syllable fusion in Cantonese connected speech*. (MPhi thesis). University of Hong Kong.

Lee, Y., & Goldrick, M. (2008). The emergence of sub-syllabic representations. *Journal of Memory and Language*, *59*, 155–168. https://doi.org/10.1016/j.jml.2008.03.002

Levelt, W. J. M. (1992). Accessing words in speech production: Stages, processes and representations. *Cognition*, *42*, 1–22.

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. https://doi.org/10.1017/s0140525x99001776

Levelt, W. J. M., Schriefers, H., Vorberg, D., Meyer, A. S., Pechmann, T., & Havinga, J. (1991). The time course of lexical access in speech production: A study of picture naming. *Psychological Review*, *98*, 122–142. https://doi.org/10.1037/0033-295x.98.1.122

Li, C., Wang, M., Kim, S. Y., Bolger, D. J., & Wright, K. (2022). Phonological preparation in Korean: Phoneme, or syllable or another unit? *Language and Speech*, *65*, 337–353. https://doi.org/10.1177/00238309211020039

Lin, W.-C., & Chen, J.-Y. (2003). Masked priming of the segmental syllable in Mandarin Chinese speech production: More evidence. In *Proceedings of the 42nd annual meeting of the Chinese Psychological Association. Taipei, Taiwan*.

Lin, Y.-H. (2007). *The sounds of Chinese*. Cambridge University Press.

Luce, P. A., & Pisoni, D. B. (1998). Recognizing spoken words: The neighborhood activation model. *Ear and Hearing*, *19*, 1–36. https://doi.org/10.1097/00003446-199802000-00001

MacKay, D. G. (1970). Spoonerisms: The structure of errors in the serial order of speech. *Neuropsychologia*, *8*, 323–350. https://doi.org/10.1016/0028-3932(70)90078-3

McCawley, J. (1978). What is a tone language? In V. Fromkin (Ed.), *Tone: A linguistic survey* (pp. 113–131). Academic Press.

Meyer, A. S. (1990). The time course of phonological encoding in language production: The encoding of successive syllables of a word. *Journal of Memory and Language*, *29*, 524–545. https://doi.org/10.1016/0749-596x(90)90050-a

Meyer, A. S. (1991). The time course of phonological encoding in language production: Phonological encoding inside a syllable. *Journal of Memory and Language*, *30*, 69–89. https://doi.org/10.1016/0749-596x(91)90011-8

Meyer, A. S., & Belke, E. (2007). Word form retrieval in language production. In G. Gaskell (Ed.), *The Oxford handbook of psycholinguistics* (pp. 471–487). Oxford University Press.

Meyer, A. S., & Schriefers, H. (1991). Phonological facilitation in picture-word interference experiments: Effects of stimulus onset asynchrony and types of interfering stimuli. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *17*, 1146–1160. https://doi.org/10.1037/0278-7393.17.6.1146

Myers, J. (2010). Chinese as a natural experiment. *The Mental Lexicon*, *5*, 421–435. https://doi.org/10.1075/ml.5.3.09mye

Nakayama, M., & Saito, S. (2014). Within-word serial order control: Adjacent mora exchange and serial position effects in repeated single-word production. *Cognition*, *131*, 415–430. https://doi.org/10.1016/j.cognition.2014.03.002

Neergaard, K. D., & Huang, C.-R. (2019). Constructing the Mandarin phonological network: Novel syllable inventory used to identify schematic segmentation. *Complexity*, *2019*, 1–21. https://doi.org/10.1155/2019/6979830

Neergaard, K. D., Xu, H., & Huang, C.-R. (2016). Database of Mandarin neighborhood statistics. In *Proceedings of the 10th international conference on language resources and evaluation, REC 2016* (pp. 4032–4036).

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). Mouton.

Nozari, N., & Dell, G. S. (2012). Feature migration in time: Reflection of selective attention on speech errors. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *38*, 1084–1090. https://doi.org/10.1037/a0026933

O'Séaghdha, P. G. (2015). Across the great divide: Proximate units at the lexical-phonological interface. *Japanese Psychological Research*, *57*, 4–21. https://doi.org/10.1111/jpr.12074

O'Séaghdha, P. G., & Chen, J.-Y. (2009). Toward a language-general account of word production: The proximate units principle. In N. A. Taatgen & H. van Rijn (Eds.), *Proceedings of the 31st annual conference of the Cognitive Science Society* (pp. 68–73). Cognitive Science Society.

O'Séaghdha, P. G., Chen, J.-Y., & Chen, T.-M. (2010). Proximate units in word production: Phonological encoding begins with syllables in Mandarin Chinese but with segments in English. *Cognition*, *115*, 282–302. https://doi.org/10.1016/j.cognition.2010.01.001

O'Séaghdha, P. G., & Marin, J. W. (2000). Phonological competition and cooperation in form-related priming: Sequential and nonsequential processes in word production. *Journal of Experimental Psychology: Human Perception and Performance*, *26*, 57–73. https://doi.org/10.1037/0096-1523.26.1.57

Otake, T., Hatano, G., Cutler, A., & Mehler, J. (1993). Mora or syllable? Speech segmentation in Japanese. *Journal of Memory and Language*, *32*, 258–278. https://doi.org/10.1006/jmla.1993.1014

Pérez, E., Santiago, J., Palma, A., & O'Séaghdha, P. G. (2007). Perceptual bias in speech error data collection: Insights from Spanish speech errors. *Journal of Psycholinguistic Research*, *36*, 207–235. https://doi.org/10.1007/s10936-006-9042-7

Pinet, S., Dell, G. S., & Alario, F.-X. (2019). Tracking keystroke sequences at the cortical level reveals the dynamics of serial order production. *Journal of Cognitive Neuroscience*, *31*, 1030–1043. https://doi.org/10.1162/jocn_a_01401

Ramoo, D., Romani, C., & Olson, A. (2021). Lexeme and speech syllables in English and Hindi. A case for syllable structure. In G. Sharma & J. J. Lowe (Eds.), *Trends in South Asian linguistics*.

Rapp, B., & Goldrick, M. (2000). Discreteness and interactivity in spoken word production. *Psychological Review*, *107*, 460–499. https://doi.org/10.1037/0033-295x.107.3.460

Repetti, L. (1991). A moraic analysis of raddopiamento fonosintattico. *Rivista di Linguistica*, *3*, 307–330.

Roelofs, A. (1999). Phonological segments and features as planning units in speech production. *Language and Cognitive Processes*, *14*, 173–200. https://doi.org/10.1080/016909699386338

Roelofs, A., & Meyer, A. S. (1998). Metrical structure and planning in production of spoken words. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *24*, 922–939. https://doi.org/10.1037/0278-7393.24.4.922

Romani, C., Galluzzi, C., Bureca, I., & Olson, A. (2011). Effects of syllable structure in aphasic errors: Implications for a new model of speech production. *Cognitive Psychology*, *62*, 151–192. https://doi.org/10.1016/j.cogpsych.2010.08.001

Sadat, J., Martin, C., Costa, A., & Alario, F.-X. (2014). Reconciling phonological neighborhood effects in speech production through single trial analysis. *Cognitive Psychology*, *68*, 33–58. https://doi.org/10.1016/j.cogpsych.2013.10.001

Saito, A., & Inoue, T. (2017). The frame constraint on experimentally elicited speech errors in Japanese. *Journal of Psychological Research*, *46*, 583–596. https://doi.org/10.1007/s10936-016-9454-y

Santiago, J., Pérez, E., Palma, A., & Stemberger, J. P. (2007). Syllable, word, and phoneme frequency effects in Spanish phonological speech errors: The David effect on the source of the error. In C. Schütze & V. Ferreira (Eds.), *MIT working papers in linguistics 53. The state of the art in speech error research: Proceedings of the LSA Institute Workshop*. MIT.

Schiller, N. O. (1997). The role of the syllable in speech production. Evidence from lexical statistics, metalinguistics, masked priming, and electromagnetic midsagittal articulography. (Doctoral dissertation).

Schiller, N. O. (1998). The effect of visually masked syllable primes on the naming latencies of words and pictures. *Journal of Memory and Language*, *39*, 484–507. https://doi.org/10.1006/jmla.1998.2577

Schiller, N. O. (2000). Single word production in English: The role of subsyllabic units during phonological encoding. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *26*, 512–528. https://doi.org/10.1037/0278-7393.26.2.512

Schiller, N. O., Meyer, A. S., Baayan, R. H., & Levelt, W. J. (1996). A comparison of lexeme and speech syllables in Dutch. *Journal of Quantitative Linguistics*, *3*, 8–28. https://doi.org/10.1080/09296179608590060

Sevald, C., Dell, G. S., & Cole, J. (1995). Syllable structure in speech production: Are syllables chunks or schemas? *Journal of Memory and Language*, *34*, 807–820. https://doi.org/10.1006/jmla.1995.1035

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). 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). 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. https://doi.org/10.1017/s0952675700000609

Shattuck-Hufnagel, S. (1987). The role of word onset consonants in speech production planning: New evidence from speech error patterns. In E. Keller & M. Gopnik (Eds.), *Motor and sensory processes of language* (pp. 17–51). Erlbaum.

Shattuck-Hufnagel, S. (1992). The role of word structure in segmental serial ordering. *Cognition*, *42*, 213–259. https://doi.org/10.1016/0010-0277(92)90044-i

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. https://doi.org/10.1016/s0022-5371(79)90554-1

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. https://doi.org/10.1111/cogs.12047

Starreveld, P. A., & La Heij, W. (1996). Time-course analysis of semantic and orthographic context effects in picture naming. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *22*, 896–918. https://doi.org/10.1037/0278-7393.22.4.896

Stemberger, J. P. (1982/1985). *The lexicon in a model of language production*. Garland.

Stemberger, J. P. (1983). *Speech errors and theoretical phonology: A review*. Distributed by IULC.

Stemberger, J. P. (1990). Word shape errors in language production. *Cognition*, *35*, 123–157. https://doi.org/10.1016/0010-0277(90)90012-9

Tamaoka, K., & Makioka, S. (2009). Japanese mental syllabary and effects of mora, syllable, bi-mora and word frequencies on Japanese speech production. *Language and Speech*, *52*, 79–112. https://doi.org/10.1177/0023830908099884

Terao, Y. (2022). The value of spontaneous speech errors as basic data in language production research. *Cognitive Studies: Bulletin of the Japanese Cognitive Science Society*, *29*, 74–84.

Verdonschot, R. G., Han, J.-I., & Kinoshita, S. (2021). The proximate unit in Korean speech production: Phoneme or syllable? *Quarterly Journal of Experimental Psychology*, *74*, 187–198. https://doi.org/10.1177/1747021820950239

Verdonschot, R. G., Phương, H. T. L., & Tamaoka, K. (2022). Phonological encoding in Vietnamese: An experimental investigation. *Quarterly Journal of Experimental Psychology*, *75*, 1355–1366. https://doi.org/10.1177/17470218211053244

Vitevitch, M. S. (1997). The neighborhood characteristics of malapropisms. *Language and Speech*, *40*, 211–228. https://doi.org/10.1177/002383099704000301

Vitevitch, M. S. (2002). The influence of phonological similarity neighborhoods on speech production. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *28*, 735–747. https://doi.org/10.1037/0278-7393.28.4.735

Vitevitch, M. S., & Castro, N. (2015). Using network science in the language sciences and clinic. *International Journal of Speech-Language Pathology*, *17*, 13–25. https://doi.org/10.3109/17549507.2014.987819

Vitevitch, M. S., & Luce, P. A. (1998). When words compete: Levels of processing in perception of spoken words. *Psychological Science*, *9*, 325–329. https://doi.org/10.1111/1467-9280.00064

Vitevitch, M. S., & Stamer, M. K. (2006). The curious case of competition in Spanish speech production. *Language and Cognitive Processes*, *21*, 760–770. https://doi.org/10.1080/01690960500287196

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. https://doi.org/10.1006/cogp.2000.0739

Wang, J., Wong, A. W.-K., & Chen, C.-S. (2018). Time course of syllabic and sub-syllabic processing in Mandarin word production: Evidence from the picture-word interference paradigm. *Psychonomic Bulletin & Review*, *25*, 1147–1152. https://doi.org/10.3758/s13423-017-1325-5

Wan, I.-P. (1997). The status of prenuclear glides in Mandarin Chinese: Evidence from speech errors. *Chicago Linguistics Society*, *33*, 417–428.

Wilshire, C. E. (1998). Serial order in phonological encoding: An exploration of the "word onset effect" using laboratory-induced errors. *Cognition*, *68*, 143–166. https://doi.org/10.1016/s0010-0277(98)00045-6

Wilshire, C. E., & McCarthy, R. A. (1996). Experimental investigations of an impairment in phonological encoding. *Cognitive Neuropsychology*, *13*, 1058–1098. https://doi.org/10.1080/026432996381782

Winskel, H., & Ratitamkul, T. (2020). The initial functional unit when naming words and pseudowords in Thai: Evidence from masked priming. *Journal of Psychological Research*, *49*, 275–290. https://doi.org/10.1007/s10936-020-09687-7

Wong, A. W.-K., & Chen, H.-C. (2008). Processing segmental and prosodic information in Cantonese word production. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *34*, 1172–1190. https://doi.org/10.1037/a0013000

Wong, A. W.-K., & Chen, H.-C. (2009). What are effective phonological units in Cantonese spoken word planning? *Psychonomic Bulletin & Review*, *16*, 888–892. https://doi.org/10.3758/PBR.16.5.888

Wong, A. W.-K., & Chen, H.-C. (2015). Processing segmental and prosodic information in spoken word planning: Further evidence from Cantonese Chinese. *Japanese Psychological Research*, *57*, 69–80. https://doi.org/10.1111/jpr.12054

Wong, W. Y. P. (2006). *Syllable fusion in Hong Kong Cantonese connected speech*. (Doctoral dissertation). The Ohio State University.

AUTHOR BIOGRAPHIES

John Alderete is a professor of Linguistics and Cognitive Science at Simon Fraser University in Canada.

Padraig G. O'Seaghdha is a professor of Psychology and the acting director of Cognitive Science at Lehigh University in the United States.

How to cite this article: Alderete, J., & O'Séaghdha, P. G. (2022). Language generality in phonological encoding: Moving beyond Indo-European languages. *Language & Linguistics Compass*, e12469. https://doi.org/10.1111/lnc3.12469