Jason Shaw and Shigeto Kawahara*

Predictability and phonology: past, present and future

Abstract: Many papers in this special issue grew out of the talks given at the Symposium "The role of predictability in shaping human language sound patterns," held at Western Sydney University (Dec. 10–11, 2016). Some papers were submitted in response to an open call; others were invited contributions. This introduction aims to contextualize the papers in the special issue within a broader theoretical context, focusing on what it means for phonological theory to incorporate gradient predictability, what questions arise as a consequence, and how the papers in this issue address these questions.

1 Predictability in the generative enterprise

Predictability has always been central to understanding sound patterns in human language. The modern theoretical landscape features two kinds of predictability: (1) the general notion of probability, which we will refer to as gradient predictability, and (2) the theory-specific notion of predictability that is dichotomous, as developed in early generative phonology.

As an example of gradient predictability, Hayes and Wilson (2008) relate grammatical well-formedness to the probability of phoneme sequences in representative corpora. Chomsky (1957) argues that this notion of predictability is unrelated to grammatical well-formedness. The celebrated arguments come from the domain of syntax; for example, the transitional probability from *fragile* to *whale* and from *fragile* to *of* are the same ($=0$), but only the latter is deemed ungrammatical (p. 16). These arguments have since been challenged (Pereira 2000), but Chomsky's conclusions determined the trajectory of phonological theory (Chomsky and Halle 1968). Most pertinently, these arguments afforded an alternative notion of predictability in which predictability is dichotomous in nature and differentiates only predictable patterns, i.e. $P = 1$, from unpredictable patterns, i.e. $P < 1$ (although in practice exceptions to deterministic rules were tolerated). For example, post-tonic coronal stops in English become predictably flaps (i.e. $P([\text{-cont}]|\hat{V}[\text{cor}]V) = 0$ and $P([\text{ɾ}]|\hat{V}[\text{cor}]V) = 1$), and on this basis, it was posited that English has a rule of flapping. Under this dichotomous conceptualization of predictability, phonological features could be either "predictable" from their phonological context, in which case they are derived by rule, or "unpredictable", in which case they need to be stored in the lexicon.

From a historical perspective, the dichotomous notion of predictability did not arise for want of appropriate mathematical tools. Before the advent of generative grammar, Information Theory offered a set of formal tools for expressing information in terms of gradient predictability (Shannon and Weaver 1949). Information is defined probabilistically in terms of the amount of uncertainty (Entropy) associated with a message, encoded as sequences of symbols. A random variable, x, in a set of N elements has an Entropy, which is a function of the number of members in N and their probabilities. By giving a definite value to x, we remove Entropy and communicate information. Entropy is defined as average predictability within

*Corresponding author: Shigeto Kawahara, The Institute of Cultural and Linguistic Studies, Keio University, 2-15-45 Mita, Tokyo 108-8345, Japan, E-mail: kawahara@icl.keio.ac.jp
Jason Shaw: Department of Linguistics, Yale University, 370 Temple Street, New Haven, CT 06520, USA, E-mail: jason.shaw@yale.edu

a given context, predictability itself being defined as Surprisal, which is the negative log of contextual probability.

$$\text{Contextual probability } = p(x|\text{Context}) \tag{1}$$

$$\text{Surprisal } = -\log_2 p(x|\text{Context}) \tag{2}$$

$$\text{Entropy } = -\sum_{i=1}^{N} p(x_i|\text{Context}) \log_2 p(x_i|\text{Context}) \tag{3}$$

Zipf's (1949) work on frequency effects, including the finding that word length is inversely proportional to word frequency, inspired the application of Information Theory to phonology (Cherry et al. 1952; Hockett 1967). As Yang (2008: 206) points out, the conception of syntactic analysis laid out in Chomsky (1955) also had a direct information-theoretic interpretation and even raised the consideration that "statistical considerations" may be relevant to grammaticality. However, Chomsky (1957: 19) cites Shannon and Weaver (1949) in the context of a critical examination of Finite State Grammar as a model of English syntax, and ultimately rejects it. In the domain of phonology, Chomsky and Halle (1968: 110) relegate effects of gradient predictability, e.g. effects of word frequency on duration or vowel reduction, to "performance," which they locate outside the scope of phonological "competence". This decision substantially diminished the role of gradient predictability in phonological theory.

Besides locating known effects of gradient predictability outside of the grammar, the pursuit of deterministic models of competence also directed phonological inquiry away from modeling large corpora. The assumption was that corpora cannot reveal insights delineating *possible* words from actual, existing words. Halle (1978) drew a line between *brick* and *blick* vs. *bnick* – a line between possible forms and impossible forms, rather than attested forms and unattested forms. What mattered in generative phonology was which forms native speakers of a language find grammatical, not whether particular words exist or not. Limiting data to dichotomous human judgements, grammatical vs. ungrammatical, may have curtailed the discovery of probabilistic phonological knowledge. For a time, probabilistic dependencies inferred from the lexicon played a very minor role, if any, in phonological theorizing, a practice which has been criticized (Ohala 1986).

2 From dichotomous to gradient predictability

The contemporary coexistence of the classic notion of dichotomous predictability with the more general notion of gradient predictability follows from the convergence of several intellectual paths on probabilistic models of phonology. One involves a shift in the empirical base of theory development towards experimental data, including systematic elicitation of speaker judgements, and towards larger datasets more generally. As Pierrehumbert (2001) pointed out, probabilistic models are natural companions to fields that use experimental or corpus data. This is not necessarily because the systems under study are inherently stochastic (they may be) but, rather, because the data are noisy. A probabilistic model is needed to relate a deterministic theory to noisy data.

There has been a general increase in the size of the data sets typically considered to be representative of a phonological pattern. One development contributing to this trend was an interest in formal learning algorithms (Boersma and Hayes 2001; Goldwater and Johnson 2003; Hayes and Wilson 2008) energized by Optimality Theory (Prince and Smolensky 1993/2004). With interest in formal learning algorithms in phonology came questions about valid empirical tests, including how to approximate the linguistic input of a learner. The challenge of learning phonological patterns, even in the presence of exceptions (Boersma and Hayes 2001), benefits from a realistically large data set. Following Pierrehumbert (2001), then, the shift to probabilistic models follows naturally from the trend to inform phonological theory with experimental and corpus data.

Another path towards probabilistic models has been the body of evidence indicating that phonological grammar itself is deeply probabilistic: phonological patterns tend to generalize according to their probability in the lexicon (Albright and Hayes 2003; Ernestus and Baayen 2003; Hayes and Londe 2006). Importantly, this includes general rules as well as minority patterns, which might have been treated as "exceptions" under a dichotomous notion of predictability (Zuraw 2000).

Third, analysis of phonetic data has improved to the point that it is now possible to differentiate between gradient phonetic reduction and variable alternation between categorical phonological forms. Improved methods have revealed cases in which what looks superficially like phonetic reduction should instead be treated as categorical phonological variation (Shaw and Kawahara 2018). The degree to which probabilistic effects on phonetic reduction can be relegated to extra-grammatical performance has likewise been challenged. The same probabilistic factors that influence phonetic reduction, arguably an aspect of performance, also influence phonological processes such as segment deletion, which have always been under the purview of competence (Cohen Priva 2015).

As the field now turns increasingly to probabilistic models, it seems natural to ask how closely the architecture of our models should resemble those developed at first by setting aside gradience. It is notable that many approaches can be characterized as adding stochastic components to (once) deterministic models; these include probabilistic rules (Labov 1969; Sankoff and Labov 1979; Albright and Hayes 2003), Stochastic OT (Boersma and Hayes 2001), and Noisy Harmonic Grammar (Coetzee and Kawahara 2013). Other models are probabilistic in the sense that they maximize the probability of a surface form given an underlying representation (Goldwater and Johnson 2003; Jarosz 2006). In this case, the conditional probability of a surface form given an underlying form need not resemble the probability of the surface form in any corpus.

Other approaches represent more directly the probability of words in corpora. The stochastic phonology of Coleman and Pierrehumbert (1997) represents the well-formedness of words as joint probabilities over phonological constituents: syllable onsets and rimes specified for word position as well (see also Frisch et al. 2000). The approach of Hayes and Wilson (2008) is conceptually similar in that well-formedness is based only on the surface probability of forms. Other approaches deploying Information Theory (Shannon and Weaver 1949) in some capacity as well rely only on surface representations (Hume and Bromberg 2005; Hall 2009; Goldsmith and Riggle 2012; Hume and Mailhot 2013; Cohen Priva 2015). Which of these approaches is more appropriate is a potentially empirical question.

One possibility is that probabilistic patterns are of precisely the same type as deterministic (i.e. $P = 1$) patterns in that they draw from the same set of possible phonological targets and conditioning contexts. If this is found to be the case, then extending models of deterministic patterns by adding probabilities seems appropriate. A key advantage of this approach is that the restrictiveness of theories developed to capture possible words can be brought to bear on probabilistic phonological patterns. However, the models may be too restrictive. Gradient patterns may include new types of interactions between phonological units or new conditioning contexts that are impossible to express insightfully in existing frameworks. In this case, phonological theory may be better served by analytical tools developed specifically for probabilistic patterns.[1]

3 Characterizing gradient predictability

3.1 Contexts for predictability

3.1.1 Phonological contexts

The precise range of phonological contexts that enter into the computation of gradient predictability is an open question. Local phonological contexts, covering the immediately adjacent (preceding/following) segments, or non-local contexts that can be construed as local by either considering articulation (Gafos 1999) or

---

1 We would like to thank Dustin Bowers for productive discussion of this point.

by projecting relevant features to tiers (Hayes and Wilson 2008) retain their relevance in probabilistic phonology. However, there is evidence for non-local effects as well; for example, Albright and Hayes (2003) show that probabilistic rules that span several segments contiguous with the target of the rule outperform a more general local rule in predicting the form of the English past tense. In addition, the rules enforcing the contiguity requirement provide a better match to human behavior than rules that relax this requirement.

In predicting segment duration and deletion rates in English, Cohen Priva (2015) uses a measure of predictability that takes into account all preceding segments in a word. A key result is that the average predictability of a segment across the words of a lexicon ("informativity") contributes to segment duration and deletion likelihood, independently of segment predictability in a particular word.[2] These approaches incorporate gradient effects of non-local segments on phonetic and phonological behavior. However, even in strictly local environments, there are open questions about which segments and features can interact.

Whang (2018) argues that variable deletion of devoiced high vowels in Japanese is conditioned locally by the predictability of the vowel given the the feature [high] and the preceding consonant. Of the two high vowels, /i/ and /u/, some preceding consonantal environments strongly predict that only one can occur. Whang (2018) argues that high vowels are more likely to be deleted when they have high contextual probability. Related measures of vowel predictability have been shown to condition vowel duration as well, in ways that interact with vowel quality and other factors (Shaw and Kawahara 2017). Although both Whang (2018) and Shaw and Kawahara (2017) acknowledge effects of the conditional probability of a vowel given the preceding consonant, they consider different sets of vowels as relevant to the computation – for the purpose of predicting deletion, only the predictability of the set of high vowels seems to be relevant; for predicting vowel duration, the entire probability distribution over vowels is required. Other results show that not all patterns of gradient predictability, even when strictly local, necessarily impact phonological behavior. Becker et al. (2011) demonstrate that in Turkish, listeners tend to ignore the height and backness of adjacent vowels when making decisions about consonant voicing, despite patterns in the lexicon that indicate consonant voicing is gradiently predictable from these vowel features (see also Hayes et al. 2009). Thus, while various measures of gradient predictability have exposed aspects of phonological knowledge hidden by the dichotomous notion of predictability, including some long-distance patterns, it is not the case that "anything goes." Not all patterns that are reflected in the lexicon and contain "information" about phonological form are necessarily deployed by language users.

3.1.2 Other contextual factors

Besides phonological context, there are other factors that influence gradient predictability. For example, for words in the lexicon, if the listener knows what word is likely to be uttered, then they will also know which phonological form is likely. Another relevant line of research has pursued the thesis that variation in phonetic and phonological form subserves effective communication of meaning (Hall et al. 2016). From this standpoint, the phonetic robustness of the signal trades off with the predictability of a particular message (c.f., Lindblom 1990). Thus, the broader syntactic, semantic, and discourse context that conditions the predictability of a particular word should also influence phonological form and phonetic detail.

Support for this conjecture comes from several studies showing gradient effects of word frequency and other measures of predictability on phonetic duration and vowel reduction (Jurafsky et al. 2001; Aylett and Turk 2004; Bell et al. 2009). Low contextual predictability seems to condition robust phonetic signaling of a word, at least in these cases (cf. Kuperman et al. 2007). There is now evidence that listeners also encode more phonetic detail when a segment has relatively low contextual predictability, which may provide the seeds to sound change (Manker 2017), particularly if embedded in cycles of internal feedback (Wedel 2007) or if a segment/word is on average more/less predictable across contexts (Seyfarth 2014; Cohen Priva 2015). Consistent

---

2 A similar result has been found at the level of words, i.e. average predictability of a word across contexts is predictive of word duration (Seyfarth 2014).

with this view, Wedel et al. (2013) found that minimal pair counts are a significant predictor of whether a phoneme undergoes merger.

Another factor that plays into phonetic predictability is talker identity. Listeners are better at recognizing words from familiar talkers than from unfamiliar talkers (Nygaard 2005), presumably because they can better predict the precise phonetic details of a word when the talker is known. Even expectations about a talker's gender (Johnson et al. 1999) or regional background (Hay and Drager 2010) can change the way that listeners parse the phonetic signal into phonological categories. The older age of a talker, as indicated through the phonetic details of the voice, facilitates processing of words perceived to be older or old-fashioned (Walker and Hay 2011). The specific location of the speaking event can also provide context that influences the phonetic details of speech and how they are perceived (Hay et al. 2017). When deprived of higher-level sources of segment predictability, including word identity, talker identity, situational context, etc., listeners show rather severe degradation in their ability to identify segments (Shaw et al. 2018). These results indicate that in speech comprehension listeners deploy knowledge relating to context, beyond just phonological context, to interpret the phonetic details of their speech experience. Parsing words from the signal may rest on the appropriate distribution of signal redundancy relative to message predictability.

3.2 Abstraction and grammar

Both the scope of gradient predictability effects within the phonology and how they might interact with extra-phonological aspects of message predictability remain open questions. At one extreme, there is the view that predictability plays *no* role (Chomsky and Halle 1968). Another view is that prosody is the means by which languages deal with predictability (Turk and Shattuck-Hufnagel 2014). For example, under the Smooth Signal Redundancy hypothesis, prosody functions to distribute information more evenly across the signal (Aylett and Turk 2004). However, this is unlikely to be the complete story, as the prosodic parse of a given utterance cannot explain why average predictability across contexts influences word duration (Seyfarth 2014) and segment duration (Cohen Priva 2015) even as prosodic structure and local moment-to-moment predictability vary. The other extreme is that predictability is the major driving force shaping phonological patterns (Hall et al. 2016). An intermediate position is to use the grammar to delineate the effects of predictability, a.k.a. "grammar-dominance". Coetzee and Kawahara (2013) argue that grammar sets the limits of what patterns of variation are possible, and all that frequency can do is to determine how variation is realized within these limits.

Language users appear to be leveraging a substantial amount of available information to make predictions about speech events likely to occur in their environment at a given moment in time. Phonological factors conditioning categorical alternations constitute one of many types of patterns to which listeners and speakers show sensitivity. On the other hand, phonological patterns show a degree of conventionalization that seems to be isolated from the many degrees of freedom involved in moment-by-moment computation of expectations. To expand on this point, let us take vowel harmony as an example. Walker (2011) examines cases of vowel harmony in which a vowel feature spreads from a less prominent position to a more prominent position. For instance, the presence of a post-tonic high vowel in Central Veneto causes the stressed vowel to raise. Thus, the stressed vowel in [bév-o] 'drink (1sg pres. ind.)' contrasts with the stressed vowel in [bív-i] 'drink (2sg pres. ind.)'. She argues that harmony in these cases helps to convey the presence of a marked feature; within Hall et al.'s (2016) framework, this could be viewed as increasing signal redundancy, because cues to the [+high] feature are enhanced. We can ask, then, to what extent we should expect spontaneous spreading of features in low-predictability environments or in noisy environments in which the signal may be degraded. On the face of it, we do not expect that vowels spontaneously harmonize in, e.g. noisy environments, despite the potential benefits of redundancy. Similarly, we do not expect that vowel harmony is blocked when the communication channel is particularly clear. To the extent that this does not happen, it seems legitimate to ask why. Speech patterns do change in noisy environments (Brumm and Zollinger 2011); however, the particular ways in which speech changes in response to environmental noise seem to differ from the ways in which speech changes in response to predictability (for an explicit comparison of these factors; see also, e.g. Zhao

and Jurafsky 2009). It may be that background noise does not induce vowel harmony for the same reason that vowel height does not induce stop voicing alternations in Turkish (Becker et al. 2011).

Thus, while we may need to broaden the scope of factors that are relevant to contextual predictability beyond those of the most localist models of phonology, the primary objective remains the same – identifying the dimensions that characterize knowledge of sound patterns. In this endeavor, there is contiguity in phonology across conceptions of predictability as dichotomous and as gradient. It also seems clear that the network of conditional dependencies that defines phonological knowledge cannot be read directly off the lexicon or frequencies in corpora, although they provide one useful angle. Knowledge of sound patterns features both blind spots, i.e. a lack of sensitivity to some conditional relationships attested in corpora (Becker et al. 2011), and hallucinations, as we may perceive forms that are likely even in the absence of phonetic evidence (Wilson 2016, see also Dupoux et al. 1999). These are good reasons to model grammatical knowledge as the set of *possible* forms, rather than existing words (Halle 1978). Padgett (2003) makes this point very clear:

> The idea of neutralization avoidance, if understood in the wrong way, can make strange predictions. For example, consider the fact that Standard English has the words *beat* [bit], *boot* [but], and *peat* [pit], but no *poot* [put]...[I]f there were a process backing [i] to [u], would we expect that it might affect [pit] but not [bit], since only the latter would entail a neutralization (with [but])?...These questions arise when we take the domain of explanation to be the set of actual lexical items in a language. But this is in fact not the practice in generative phonology. Instead, theories model the set of *possible* words of a language.... (pp. 78–79; emphasis in the original)

Even in the absence of an actual word, phonological knowledge seems to "reserve space" for additional possible (or likely) words, avoiding the collapse of categories. In this sense, phonological systems are optimized for lexical expansion in a way that can only be explained by abstract phonological structure, a point also made by Pierrehumbert (2016).

To summarize, there are a number of issues that arise when we consider how gradient predictability influences sound patterns. Key points of agreement include the necessity of abstraction. Linguistic forms, whether words, syllables, segments, features, gestures, etc., play a foundational role in understanding sound patterns across the range of perspectives surveyed here. Abstract linguistic units of various degrees of granularity provide the basis for computing gradient predictability, even if they are considered alongside extra-linguistic factors such as the identity of the talker or physical location of the speech event. Another point of agreement is that the phonetic signal is systematically impacted by gradient predictability. Why this is the case, including related issues of precisely which aspects of context contribute to predictability, remains an open question.

Since the central question involving gradient predictability involves the relation between measurable phonetic properties and the abstract linguistic forms that they signify, it would be surprising if phonological theory, which ostensibly dictates this mapping, played no role in the solution. How best to leverage phonological insights obtained by, at first, abstracting away from phonetic facts and gradient predictability is in part the question that motivated us to contextualize this special issue within generative phonology. It is unclear at this point whether it will be theoretical insights from generative phonology that will be leveraged creatively to bring order to the facts of gradient predictability, or rather that analytical tools developed first to deal with gradient predictability will ultimately provide deeper theoretical insights and superior coverage of the facts. We acknowledge, of course, that these alternatives are not mutually exclusive. With that, we now turn to the papers in the special issue.

4 The current volume

The 13 papers in this volume take up a range of positions related to the issues introduced above, including the crucial issues of context, the level at which predictability effects shape sound patterns, and the theoretical status of predictability. They bring new case studies from various languages – Bardi (Babinski and Bowern 2018), Japanese (Sano 2018; Turnbull 2018), Korean (Kawahara and Lee 2018) – and from L2 populations

(Baese-Berk et al. 2018; Olejarczuk et al. 2018). They illustrate influences of predictability on a wide range of empirical phenomena, including lenition (Foulkes et al. 2018), segment duration (Clopper et al. 2018; Sano 2018), deletion (Turnbull 2018; Kawahara and Lee 2018), phonemic mergers (Babinski and Bowern 2018), spoken-word recognition (Baese-Berk et al. 2018), phonetic category learning (Olejarczuk et al. 2018), and articulatory movements (Tomaschek et al. 2018).

The papers also exhibit a range of computational modeling approaches. These include an application of factor analysis to model how phonological features structure talker-specific phonetic detail (Chodroff and Wilson 2018), a model of phonetic category acquisition based on error-driven learning (Olejarczuk et al. 2018), and several studies deploying regression models to demonstrate how predictability shapes sound patterns in speech corpora (Clopper et al. 2018; Sano 2018; Turnbull 2018) and experimental data (Baese-Berk et al. 2018), including an application of quantile regression revealing non-linear effects of word frequency on articulatory movement trajectories (Tomaschek et al. 2018).

The volume also comprises papers that make other important methodological contributions, including an examination of the interrelatedness of frequency, predictability and informativity, with implications for false positives (Cohen Priva and Jaeger 2018), explicit comparison of different estimates of predictability drawing from written corpora, spoken corpora and cloze probability (Clopper et al. 2018), and experimental results with implications for how frequency is used to approximate language experience (Olejarczuk et al. 2018).

With regards to the theoretical status of predictability, the papers in the volume expose different views including that predictability is directly related to signal specificity (Hall et al. 2018), that the relation between predictability and the phonetic signal is mediated by language experience (Baese-Berk et al. 2018), that predictability interacts with but is separate from the grammar (Kawahara and Lee 2018), that predictability is possibly just a methodological artifact (Cohen Priva and Jaeger 2018), and that predictability has both universal and language-specific aspects (Turnbull 2018).

We regards to the locus of predictability effects, Hall et al. explore message-based predictability or, more specifically, the predictability of meaning-bearing-units (MBUs) in context. This is in contrast to specific phonological loci of predictability investigated in other papers (e.g. Cohen Priva and Jaeger 2018; Sano 2018) and to Foulkes et al., who challenge the notion that "message" can be reduced to MBUs, since the phonetic signal carries other systematic meanings, in addition to truth-conditional meanings, including the identity of the talker. Also related to this critique is Chodroff and Wilson, whose model locates talker-specific phonetic differences within phonological features that generalize across segments. According to Hall et al. phonetic and phonological patterns serve to either reduce or enhance the signal associated with an MBU as a function of MBU predictability. They argue that the distribution of /t/ allophones in English follows from viewing predictability at the level of MBUs. Kawahara and Lee bring this perspective to an analysis of truncation patterns in Korean names, arguing that portions of first names are deleted just when they are predictable from context. They formalize this insight as a grammar-external factor, the "I-Map" (for Information Map), which conditions the ranking of faithfulness constraints in an OT grammar.

Viewed diachronically, message-based predictability may exert a bias towards preserving phonological contrasts that play a substantial role in differentiating meaning, indexed, e.g. by the number of minimal pairs for a given contrast. Babinski and Bowern demonstrate that minimal pair counts (as well as phoneme frequency) are significant predictors of phoneme merger in Bardi, extending the previous observation by Wedel et al. (2013) into another language family.

Sano's paper on the robustness of the Japanese singleton-geminate contrast also investigates the role of minimal pair count; in this case, the finding is that the duration difference between singletons and geminates is greater when this contrast differentiates minimal pairs. Although this finding is consistent with message-based predictability, Sano also examines phonological predictability, i.e. the predictability of phonological units based on other phonological units. He finds that contextual factors influencing the degree of singleton-geminate duration differences includes the sonority of the segments and their position in words (initial, medial, final). The robustness of the phonological length contrast tends to vary with the uncertainty of the length contrast given the phonological context. The tradeoff between singleton-geminate uncertainty and phonetic robustness in the phonetics follows from Hall et al.'s proposal, except that, in the Japanese case, the tradeoff is observed for the predictability of phonological units (length contrast) as opposed to MBUs.

Turnbull analyzes segment deletion in Japanese and English at both the word and phoneme levels. The word-level analysis is consistent with the MBU as the locus of predictability effects – factors including word frequency and lexical neighborhood density influence the number of segmental deletions per word in both languages. However, at the phoneme level, both languages also show gradient predictability effects. These effects are less clearly related to MBUs. At this level of analysis, Turnbull's study reveals language-specific influences of predictability on segment deletion.

Daland and Zuraw take up a critical discussion of the locus of predictability effects on phonetic duration. They raise the possibility that some of the local effects of predictability attested in the literature could be accidental consequences of higher-level discourse factors, such as whether a word is discourse-given. This is related to the concern raised by Foulkes et al. that prosodic structure may be correlated with predictability, yet few studies control for the influence of prosody on the phonetic signal. Daland and Zuraw also question some of the mechanistic interpretations given to predictability effects, particularly those located within the production system or the lexicon. They raise the interesting possibility that ease of perception could introduce a bias in word duration, if words that were easy to recognize in context are encoded as shorter than words that are more difficult to recognize.

Baese-Berk et al. report a perception study investigating interactions between predictability and speech rate. Their measures of predictability, lexical frequency and collocation frequency, are distinctly message-based and local. In the experiment, these measures interact with speech rate manipulations known to influence how words are parsed from the speech stream. Comparison between native and non-native listeners revealed differential effects of how speech rate interacts with lexical frequency. Non-native listeners fail to parse function words when they are preceded by high frequency lexical items at a greater rate than native speakers.

Tomaschek et al. interpret frequency in part as an index motor proficiency. They show that vowel articulation varies non-linearly with word frequency. As frequency increases from low to medium, the degree of curvature in the articulatory trajectory of the tongue is reduced, indicating a higher extrema for the vowel target. This aspect is consistent with a trade-off between word predictability and signal robustness, cf. Hall et al. However, further increases in frequency from medium to high frequency words show a return to the degree of trajectory curvature found at low frequency. This change in vowel articulation goes in the opposite direction of the predictability-signal robustness tradeoff. Tomaschek et al. (2018) interpret the result as the effect of practice on articulation – with practice, extreme articulatory targets can be achieved efficiently, i.e. without requiring more time.

Likewise, related to the interpretation of lexical frequency, Olejarczuk et al. propose that linguistic representations are neither faithful reflections of language exposure nor parametric summaries. Instead, they propose that phonetic categories are acquired through predictive learning. Accordingly, phonetic parameters that differ greatly from predicted values influence learning disproportionally. Support for the proposal comes from a distributional learning experiment, the results of which are modeled using an error-driven learning model.

Some of the papers highlight important methodological issues associated with analyzing sound patterns in terms of predictability. Clopper et al. compare several measures of predictability, including conditional (trigram) probabilities drawn from written and spoken corpora, as well as cloze probabilities in a behavioral task. Although these are all correlated measures of local predictability, they have different effects on vowel duration and, moreover, they interact in distinct ways with other factors, such as second mention reduction. Importantly, the claims made about predictability effects will be different depending on the specific measure of predictability incorporated into the study. Foulkes et al. also raise methodological concerns for corpus studies, including measurement error in the phonetics relative to the effect size typically found for predictability-related factors in regression models.

Finally, Cohen Priva and Jaeger take up a specific methodological issue related to concerns raised by Clopper et al. and Foulkes et al. Through a series of computational simulations, they investigate correlations between measures of frequency, predictability and informativity, examining how likely it is for, e.g. a frequency effect to masquerade as a predictability or informativity effect and vice versa. Amongst their results

is the notable asymmetry that frequency is likely to be a significant predictor when the true effect is one of informativity.

In closing, the special issue brings together a rich tapestry of empirical, methodological and theoretical contributions that will impact future research on the role of predictability in shaping sound patterns. We are optimistic that the challenge of appropriately formalizing gradient predictability will continue to inspire exploration of a richer set of analytical tools and empirical facts than have otherwise fallen under the purview of phonology. We view this to be a highly positive development in part because of the potential to catalyze new perspectives on other open issues, including the relation between synchrony and diachrony and how abstract phonological structure relates to the continuous phonetics.

References

Albright, A. & B. Hayes. 2003. Rules vs. analogy in English past tenses: A computational/experimental study. *Cognition* 90. 119–161.

Aylett, M. & A. Turk. 2004. The smooth signal redundancy hypothesis: A functional explanation for relationships between redundancy, prosodic prominence, and duration in spontaneous speech. *Language and Speech* 47(1). 31–56.

Babinski, S. & C. Bowern. 2018. Mergers in Bardi: Contextual probability and predictors of sound change. *Linguistics Vanguard* 4(S2).

Baese-Berk, M., T. H. Morrill & L. C. Dilley. 2018. Predictability and perception for native and non-native listeners. *Linguistics Vanguard* 4(S2).

Becker, M., N. Ketrez & A. Nevins. 2011. The surfeit of the stimulus: Analytic biases filter lexical statistics in Turkish laryngeal alternations. *Language* 88(2). 231–268.

Bell, A., J. M. Brenier, M. Gregory, C. Girand & D. Jurafsky. 2009. Predictability effects on durations of content and function words in conversational English. *Journal of Memory and Language* 60(1). 91–111.

Boersma, P. & B. Hayes. 2001. Empirical tests of the Gradual Learning Algorithm. *Linguistic Inquiry* 32. 45–86.

Brumm, H. & S. A. Zollinger. 2011. The evolution of the lombard effect: 100 years of psychoacoustic research. *Behaviour* 148(11–13). 1173–1198.

Cherry, E. C., R. Jakobson & M. Halle. 1952. Toward the logical description of languages in their phonemic aspect. *Language* 29. 34–46.

Chodroff, E. & C. Wilson. 2018. Predictability of stop consonant phonetics across talkers: Between-category and within-category dependencies among cues for place and voice. *Linguistics Vanguard* 4(S2).

Chomsky, N. 1955. *The logical structure of linguistic theory*. New York: Plenum Press. Published 1975.

Chomsky, N. 1957. *Syntactic structures*. The Hague: Mouton.

Chomsky, N. & M. Halle. 1968. *The sound pattern of English*. New York: Harper and Row.

Clopper, C. G., R. Turnball & R. S. Burdin. 2018. Assessing predictability effects in connected read speech. *Linguistics Vanguard* 4(S2).

Coetzee, A. W. & S. Kawahara. 2013. Frequency biases in phonological variation. *Natural Language and Linguistic Theory* 30(1). 47–89.

Cohen Priva, U. 2015. Informativity affects consonant duration and deletion rates. *Laboratory Phonology* 6(2). 243–278.

Cohen Priva, U. & F. T. Jaeger. 2018. The interdependence of frequency, predictability, and informativity. *Linguistics Vanguard* 4(S2).

Coleman, J. & J. Pierrehumbert. 1997. Stochastic phonological grammars and acceptability. In *Computational phonology: Third meeting of the ACL special interest group in computational phonology*, 49–56. Somerset: Association for Computational Linguistics.

Daland, R. & K. Zuraw. 2018. Loci and locality of informational effects on phonetic implementation. *Linguistics Vanguard* 4(S2).

Dupoux, E., K. Kakehi, Y. Hirose, C. Pallier & J. Mehler. 1999. Epenthetic vowels in Japanese: A perceptual illusion? *Journal of Experimental Psychology: Human Perception and Performance* 25. 1568–1578.

Ernestus, M. & H. Baayen. 2003. Predicting the unpredictable: Interpreting neutralized segments in Dutch. *Language* 79(1). 5–38.

Foulkes, P., G. Docherty, S. Shattuck-Hufnagel & V. Hughes. 2018. Three steps forward for predictability consideration of methodological robustness, indexical and prosodic factors, and replication in the laboratory. *Linguistics Vanguard* 4(S2).

Frisch, S., N. Large & D. Pisoni. 2000. Perception of wordlikeness: Effects of segment probability and length of the processing of nonwords. *Journal of Memory and Language* 42. 481–496.

Gafos, A. 1999. *The articulatory basis of locality in phonology*. New York: Garland.

Goldsmith, J. & J. Riggle. 2012. Information theoretic approaches to phonology: The case of finnish vowel harmony. *Natural Language and Linguistic Theory* 30(3). 859–896.

Goldwater, S. & M. Johnson. 2003. Learning OT constraint rankings using a maximum entropy model. *Proceedings of the Workshop on Variation within Optimality Theory* 111–120.

Hall, K. C. 2009. *A probabilistic model of phonological relationships from contrast to allophony*. Columbus, OH: Ohio State University dissertation.

Hall, K. C., E. Hume, F. T. Jaeger & A. Wedel. 2016. The message shapes phonology. Ms. UBC, University of Canterbury, University of Rochester and University of Arizona.

Hall, K. C., E. Hume, F. T. Jaeger & A. Wedel. 2018. The role of predictability in shaping phonological patterns. *Linguistics Vanguard* 4(S2).

Halle, M. 1978. Knowledge unlearned and untaught: What speakers know about the sounds of their language. In M. Halle, J. Bresnan & G. A. Miller (eds.), *Linguistic theory and psychological reality*, 294–303. Cambridge: MIT Press.

Hay, J. & K. Drager. 2010. Stuffed toys and speech perception. *Linguistics* 48(4).

Hay, J., R. Podlubny, K. Drager & M. McAuliffe. 2017. Car-talk: Location-specific speech production and perception. *Journal of Phonetics* 64. 94–109.

Hayes, B. & Z. Londe. 2006. Stochastic phonological knowledge: The case of Hungarian vowel harmony. *Phonology* 23. 59–104.

Hayes, B. & C. Wilson. 2008. A maximum entropy model of phonotactics and phonotactic learning. *Linguistic Inquiry* 39. 379–440.

Hayes, B., K. Zuraw, P. Siptár & Z. Londe. 2009. Natural and unnatural constraints in Hungarian vowel harmony. *Language* 85(4). 822–863.

Hockett, C. F. 1967. The quantification of functional load. *Word* 23. 301–320.

Hume, E. & I. Bromberg. 2005. Predicting epenthesis: An information-theoretic account. Talk presented at the 7th Annual Meeting of the French Network of Phonology, Aix-en-Provence, June 2nd–4th.

Hume, E. & F. Mailhot. 2013. The role of entropy and surprisal in phonologization and language change. In A. Yu (ed.), *Origins of sound patterns: Approaches to phonologization*, 29–47. Oxford: Oxford University Press.

Jarosz, G. 2006. *Rich lexicons and restrictive grammars: Maximum likelihood learning in Optimality Theory*. Baltimore, MD: Johns Hopkins University dissertation.

Johnson, K., E. A. Strand & M. D'Imperio. 1999. Auditory-visual integration of talker gender in vowel perception. *Journal of Phonetics* 27(4). 359–384.

Jurafsky, D., A. Bell, M. Gregory & W. Raymond. 2001. Probabilistic relations between words: Evidence from reduction in lexical production. In J. Bybee & P. Hopper (eds.), *Frequency and the emergence of linguistic structure*, 229–254. Amsterdam: John Benjamins.

Kawahara, S. & S. Lee. 2018. Truncation in message-oriented phonology: A case study using Korean vocative truncation. *Linguistics Vanguard* 4(S2).

Kuperman, V., M. Pluymaekers, M. Ernestus & H. Baayen. 2007. Morphological predictability and acoustic duration of interfixes in Dutch compounds. *Journal of the Acoustical Society of America* 121(4). 2261–2271.

Labov, W. 1969. Contraction, deletion, and inherent variability of the English copula. *Language* 45. 715–762.

Lindblom, B. 1990. Explaining phonetic variation: A sketch of the HandH theory. In W. J. Hardcastle & A. Marchal (eds.), *Speech production and speech modeling*, 403–439. Dordrecht: Kluwer.

Manker, J. 2017. Contextual predictability and phonetic attention. Talk given at LSA 2017, Austin, Jan 5th–8th.

Nygaard, L. C. 2005. Perceptual integration of linguistic and nonlinguistic properties of speech. In D. Pisoni & R. Remez (eds.), *The handbook of speech perception*, 390–413. Oxford: Blackwell.

Ohala, J. J. 1986. Consumer's guide to evidence in phonology. *Phonology* 3. 3–26.

Olejarczuk, P., V. Kapatsinski & H. Baayen. 2018. Distributional learning is error-driven: The role of surprise in the acquisition of phonetic categories. *Linguistics Vanguard* 4(S2).

Padgett, J. 2003. The emergence of contrastive palatalization in Russian. In E. Holt (ed.), *Optimality Theory and language change*, 307–335. Dordrecht: Kluwer Academic Press.

Pereira, F. 2000. Formal grammar and information theory: Together again. *Philosophical Transaction of Royal Society* 358. 1239–1253.

Pierrehumbert, J. B. 2001. Stochastic phonology. *GLOT* 5. 1–13.

Pierrehumbert, J. B. 2016. Phonological representation: Beyond abstract versus episodic. *Annual Review of Linguistics* 2. 33–52.

Prince, A. & P. Smolensky. 1993/2004. *Optimality Theory: Constraint interaction in generative grammar*. Malden and Oxford: Blackwell.

Sankoff, G. & W. Labov. 1979. On the use of variable rules. *Language in Society* 8. 189–222.

Sano, S. 2018. Durational contrast in gemination and informativity. *Linguistics Vanguard* 4(S2).

Seyfarth, S. 2014. Word informativity influences acoustic duration: Effects of contextual predictability on lexical representation. *Cognition* 133. 140–155.

Shannon, C. & W. Weaver. 1949. *The mathematical theory of communication*. Urbana, IL: University of Illinois Press.

Shaw, J. A., C. T. Best, G. Docherty, P. Evans, P. Foulkes, J. Hay & K. Mulak. 2018. Resilience of english vowel perception across regional accent variation. *Laboratory Phonology*.

Shaw, J. & S. Kawahara. 2017. Effects of surprisal and entropy on vowel duration in Japanese. *Language and Speech*.

Shaw, J. & S. Kawahara. 2018. Assessing surface phonological specification through simulation and classification of phonetic trajectories. *Phonology* 35.

Tomaschek, F., B. Tucker, M. Fasiolo & H. Baayen. 2018. Practice makes perfect: The consequences of lexical proficiency for articulation. *Linguistics Vanguard* 4(S2).

Turk, A. & S. Shattuck-Hufnagel. 2014. Timing is talking: What is it used for, and how is it controlled? *Philosophical Transactions of the Royal Society B*. 369(1658). 20130395. doi: 10.1098/rstb.2013.0395.

Turnbull, R. 2018. Patterns of probabilistic segment deletion/reduction in English and Japanese. *Linguistics Vanguard* 4(S2).

Walker, R. 2011. *Vowel patterns in language*. Cambridge: Cambridge University Press.

Walker, A. & J. Hay. 2011. Congruence between 'word age' and 'voice age' facilitates lexical access. *Laboratory Phonology* 2(1). 219–237.

Wedel, A. 2007. Feedback and regularity in the lexicon. *Phonology* 24(1). 147–185.

Wedel, A., S. Jackson & A. Kaplan. 2013. Functional load and the lexicon: Evidence that syntactic category and frequency relationships in minimal lemma pairs predict the loss of phoneme contrasts. *Language and Speech* 56(3). 395–417.

Whang, J. 2018. Recoverability-driven coarticulation: Acoustic evidence from Japanese high vowel devoicing. *Journal of the Acoustical Society of America* 143. 1159–1172.

Wilson, C. 2016. *Lexical statistics determine the choice of epenthetic vowel in Japanese loanword adaptation*. New Haven, CN: Yale University BA Thesis.

Yang, C. 2008. The great number crunch. *Journal of Linguistics* 44(1). 205–228.

Zhao, Y. & D. Jurafsky. 2009. The effect of lexical frequency and Lombard reflex on tone hyperarticulation. *Journal of Phonetics* 37(2). 231–247.

Zipf, G. K. 1949. *Human behavior and the principle of least effort*. Cambridge, MA: Addison-Wesley Press.

Zuraw, K. 2000. *Patterned exceptions in phonology*. Los Angeles, CA: University of California, Los Angeles dissertation.