Phonology **34** (2017) 211–219. © Cambridge University Press 2017
doi:10.1017/S0952675717000112

*Computational phonology today**

Jeffrey Heinz
University of Delaware

William J. Idsardi
University of Maryland

1 Introduction

This thematic issue almost did not happen. One of us (JH) was almost killed two days after the deadline for article submissions. As a pedestrian on a sidewalk minding his own business, he was struck by a car that ran a red light and lost control after a collision. So when we write that we are delighted to be writing this introduction, over one year later, we both really mean it.

Broadly speaking, computational phonology encompasses a variety of techniques and goals (see Daland 2014 for a survey). In this introduction we would like to highlight three aspects of current work in computational phonology: data science and model comparison, modelling phonological phenomena using computational simulations, and characterising the computational nature of phonological patterning with theorems and proofs. Papers in this thematic issue illustrate all three of these trends, and sometimes more than one of them. The way we group them in this introduction is meant to highlight the similarities between them, and not to diminish the importance of their other contributions. As we discuss these areas, we also highlight important conceptual issues that we believe are often overlooked.

2 Data science and model comparison

Data science embodies a variety of statistical and computational methods for extracting knowledge from data, and model comparison refers to statistical and mathematical principles that allow one to quantify and compare the effectiveness of different models for explaining trends observed in data. Both methods have found increasing use in phonological theory, as exemplified by three of the papers in this issue.

* E-mail: HEINZ@UDEL.EDU, IDSARDI@UMD.EDU.

We would like to thank the EMTs, doctors and therapists, as well as the many reviewers for this special issue, and especially the editors of *Phonology*, Ellen Kaisse and Colin Ewen, who provided invaluable advice and assistance over the past year with tremendous grace and humour, in somewhat extraordinary circumstances.

Keane *et al.*'s paper provides a comparison of two new similarity measures for ASL fingerspelling. Similarity plays a large role in current theorising (exemplified for instance by a special issue on phonological similarity in *Lingua*; Gallagher *et al.* 2012), and this paper allows sign-language phonology to be brought into this discussion. The two metrics make different predictions, and psycholinguistic tasks were designed to elicit judgments from fluent ASL fingerspellers that could potentially favour one model over another. Using statistical methods for model comparison (AIC, BIC), Keane *et al.* show that the evidence favours the metric based on positional similarity.

The same model-comparison techniques are also employed by Shih in her contribution. Shih argues that superadditive effects (or gang effects) are best handled with conjoined constraints, even though Harmonic Grammar (HG) can in principle account for superadditivity by the way in which it adds together constraint violations to provide an overall harmony score for each candidate. The empirical basis for her study comes from an approximately 1200-word corpus of nouns which illustrate the definite–indefinite tonal alternation in Dioula d'Odienné. She compares the effectiveness of HG grammars with and without conjoined constraints using the AIC and BIC model comparison criterion, and concludes:

> we suffer a potentially significant loss of information and explanatory power if grammars are *a priori* restricted from having constraint conjunction. Instead, the necessity and viability of conjunction must be quantitatively assessed against noisy natural language data.

Jarosz compares how well different models account for the data patterns related to sonority sequencing in child-produced Polish speech. Her major finding, based on careful statistical analysis, is that this developmental corpus shows a preference for the Sonority Sequencing Principle (SSP), which is not predicted either by the segmental statistics found in the lexicon or by learning biases ('structured generalisation') as it has been instantiated in recent computational learning models. In other words, there is evidence for a universal pressure or bias for SSP, but the precise source and nature of the bias remain a mystery.

2.1 Comments on data-science approaches

A difficult question when analysing a corpus of data is to understand the roles undifferentiated data points can play. An early example which illustrates the conceptual issues comes from Hyman (1975: 20). He classifies various types of word-forms as in Table I.

Most corpus-based approaches would assume that every occurring form is well-formed, at least to some significant degree. But, as Hyman's example makes clear, this is not a logical necessity. It is logically possible that speakers store and use extragrammatical forms; perhaps English *sphere* is one such form. Admitting this possibility makes data analysis much

	occurring	non-occurring
well-formed	*brick*	*blick*
ill-formed	*sphere*	*bnick*

Table I
A taxonomy of word types (from Hyman 1975: 20).

more difficult, because there are now two types of occurring data, which are undifferentiated, and need to be labelled in some way. Essentially, the problem becomes an unsupervised learning problem of assigning these labels to the data.

This particular example resonates in recent computational modelling efforts. Hayes & Wilson (2008: 395) write: 'in constructing a learning corpus for English onsets, we must consider the status of 'exotic' onsets such as [zw] (as in *Zwieback*), [sf] (*sphere*), and [pw] (*Puerto Rico*)'. They further explain that their corpus:

> was obtained by culling all of the word-initial onsets from the online CMU Pronouncing Dictionary (http://www.speech.cs.cmu.edu/) and removing all of the onsets that we judged to be exotic. This corpus was created before any modeling was done, so we can claim not to have tailored it to get the intended results. We obtained similar, though slightly less accurate, results for a variety of 'exotic' corpora.

Evidently, Hayes & Wilson use the scare-quoted term 'exotic' to mean what Hyman expresses by labelling words like *sphere* as 'occurring but ill-formed'. One approach to the conceptual conundrum of occurring ungrammatical forms is to equate well-formedness with probability, as Hayes & Wilson do. We critically discuss this option in the next section. Finally, this issue is not specific to phonological generalisations. In syntax, it has long been observed that actually occurring proper names (for example the band 'The The') can defy otherwise robust syntactic generalisations.

3 Simulations

Several of the papers in this issue examine various phonological issues by using computational simulations to elucidate the relevant aspects.

Nazarov & Pater present simulations of a learning device which returns probabilistic stratal OT grammars from pairs of underlying and surface forms. Specifically, the constraint-based grammar they adopt is Maximum Entropy Grammar (MaxEnt). They investigate how well this learner could infer opaque generalisations in addition to the hidden structure entailed by the stratal architecture. They examine the learner's behaviour in two case studies, Southern French tensing/laxing and Canadian

English raising/flapping, and show that it is less successful learning the opaque generalisations than their transparent counterparts, unless evidence of the stratal affiliation is provided independently. This article encourages additional research on learning opacity as 'essential to test the predictions of a general advantage for transparency, and of improved learning of opacity, given evidence of stratal affiliation'.

O'Hara's article also uses MaxEnt-based grammars and learning methods. This contribution targets another long-standing thorny issue in phonological theory: the status of abstract underlying representations, and how they might be learned, observing that one common argument is that abstract URs are too difficult to learn. O'Hara runs learning simulations motivated by an [i] ~ [∅] alternation in Klamath, which he argues is best understood as deriving from underlying abstract /e/. The MaxEnt-based learner is able to infer both the constraint weights and the underlying representations of the morphemes, including abstract /e/. O'Hara concludes that 'the learnability argument against abstract URs is not sufficient'. Perhaps most interestingly, he further argues that 'the same properties that an analyst might look for when picking an abstract UR for an alternation – feature economy, symmetry, minimising lexical gaps – are in fact emergent biases in a MaxEnt learning framework'.

Bradfield's contribution takes a broader, more philosophical perspective, asking what we can learn from simulations, and how we can improve the simulations we perform. In many respects the issues raised are not novel, but they are worth repeating, especially at this moment in phonological theory, when computational simulations are in more frequent use. The basic issue is: 'What can one conclude from a simulation?'. As Bradfield puts it, 'when building simulations, there are many choices to be made … , and the effect of [these] may not be clear'. Because there are often so many moving parts, he argues that simulations ought to be 'designed with careful analysis of the underlying theories, analyses of the sources of error, and the rest of the apparatus usual in physical and engineering science simulation studies', and that they should be 'conducted over a wide range of possible configurations and parameter settings' in order to draw reliable conclusions. These questions are examined in some detail, as Bradfield replicates and extends earlier simulations on inferring vowel systems from acoustic data. He concludes that simulations can do no more than show that the associated theory *may* be right.

3.1 Comments on simulation-based approaches

Clearly, a common theme among these papers is the use of MaxEnt simulations (Goldwater & Johnson 2003), a testament to its current influence in the field. We take this opportunity to address what we see as a very common mistake in its use: the equation of probability with grammaticality. In the article that popularised the use of MaxEnt grammars, Hayes & Wilson (2008: 383) say that 'the core idea in the application of maxent grammars to phonotactics is that well-formedness can be interpreted as

probability'. An inevitable consequence of this idea is that only finitely many words will have a value above any given threshold. This is because infinite sums are convergent only under conditions of rapidly decreasing values which must at some point fall below the threshold. So if one believes there is an ungrammatical form, it will have some non-zero probability, and only finitely many forms can have a higher probability.

To explain, in a CV language the ill-formed word shape *bap* will have some presumably non-zero value ε, and the word *ba* will have some greater probability, but it will still be less than one. There are also infinitely many other words with CV syllables only that have some non-zero probability value. As these words get longer, the probabilities must decrease, to ensure a proper probability distribution where the total probability mass equals one (that is, the sum of the probabilities must add up to one). Thus there will also be some number n such that $(CV)^n$ (i.e. a word with n CV syllables) will have a score less than ε. Simply put, longer words are inevitably less probable.

Consequently, the equation of probability with well-formedness retreats from the traditional competence/performance distinction, which treats length and grammaticality as separate factors affecting linguistic performance. The competence/performance distinction is not purely a conceptual issue; Savitch (1993) shows concretely how grammars which admit infinitely many well-formed representations can be more perspicuous.

A weaker, and more plausible, version of this idea is that there is a lawful transformation from probabilities to well-formedness scores. Lau *et al*. (2016) pursue this idea in the context of grammaticality and acceptability of sentences. They recognise that probability values cannot be directly interpretable as well-formedness, propose some methods for relating them and evaluate these methods on judgements collected experimentally. We judge that the arguments made there are valid in phonology as well.

4 The computational nature of phonology: theorems and proofs

A third way of contributing to our understanding of phonological systems is to examine their computational nature. This approach applies tools from theoretical computer science and formal language theory to the task of understanding the kinds of representations and the computational power needed to manipulate those representations in order to express linguistic generalisations.

Jardine's contribution examines surface tonal patterns in a variety of languages. He shows that well-formed tonal patterns in these languages can be described with a set of inviolable constraints over autosegmental representations. What makes this result particularly striking is the fact that every constraint identifies some connected chunk of autosegmental structure as illicit. Such constraints are STRICTLY LOCAL in a well-defined computational sense (Rogers & Pullum 2011), and the well-formed structures

they define necessarily inherit this property. Several consequences follow, of which we highlight two. Well-formedness patterns that require global inspection of the autosegmental representations are predicted not to exist. Second, while these local constraints over autosegmental representations are language-specific, they can be learned (Jardine 2016).

Graf also studies the nature of constraints and locality on surface phonotactic patterns. Building on a computational taxonomy of constraints (Heinz to appear), Graf unifies local, long-distance and tier-based generalisations by incorporating phonological domains into this taxonomy. Once the string is parsed into a particular kind of subdomain, the differences between the generalisations collapses. As in Jardine's article, the resulting constraints are inviolable, language-specific and local within each subdomain, and thereby some learnability results follow.

Both Jardine's and Graf's contributions acknowledge that phonological generalisations are finite-state. That is, 'they can be recognized using an amount of memory that is independent of the length of the input string' (Rogers 1998: 8). Hulden shows how the finite-state characterisation of phonological generalisations provides a variety of methods to verify grammatical analyses and exactly characterise classes of 'problematic' forms. His foma-toolkit (Hulden 2009) is an open-source, industrial strength reimplementation and extension of the Xerox xfst software (Beesley & Karttunen 2003) that allows researchers to easily incorporate these methods into their analyses, regardless of whether the grammatical framework is constraint-based or rule-based. As he explains, 'the methods are illustrated by practical case studies that are intended to both resolve concrete issues and be representative of typical techniques and results'. In this way, this paper is a contribution to both theory and analytic methods.

4.1 Comments on theorems and proofs

One of the common themes in the Jardine, Graf and Hulden contributions is the role that theorems and proofs play. Their methods and results are unassailable in a particular sense: if one accepts the premises, one must accept the conclusions.

We take the opportunity in this introduction to provide a very brief consumer's guide to various kinds of proofs in computational linguistics. One observation is that proofs often do not tell you exactly what you wanted to know. This almost always has to do with the premises. It is incredibly difficult to match the premises exactly with the real-world problem.

The analyst often deliberately abstracts in such a way that proofs can be constructed using known methods (for example, by reduction to another problem of known complexity). Speaking informally, the resulting models often characterise only a subset of the real-world problem or a superset of the problem, as shown in Table II.

Beginning with the upper left cell (which shows that a subset model is tractable), an example of such a demonstration is the Constraint Demotion algorithm (Tesar & Smolensky 2000), which provably learns a

	subset model	superset model
positive computational result	Recursive Constraint Demotion	finite-state interpretation of *SPE*-style rules
negative computational result	primitive OT is NP-hard	regular relations are not identifiable in the limit

Table II
A taxonomy of computational results in phonology.

constraint ranking from positive data. The premise here, though, is that the grammar is monostratal, and all constraints are given in advance (i.e. there is no online constraint conjunction). But strata and constraint conjunction are widely adopted by real-world analysts, including those in this special issue (Nazarov & Pater, Shih). There are two obvious options: give priority to the computational learning result and try to carry out phonology without strata or online constraint conjunction, or accept the real-word analyses and try to prove a new learning result with strata and/or online constraint conjunction.

The other cells can be interpreted in a similar fashion. Results falling in the cell in the upper right corner, such as Kaplan & Kay's (1994) demonstration that *SPE*-style rule systems are finite-state, offer reassurance to the real-world analysts that models built on this framework (e.g. foma) will remain tractable. For the lower left cell, Eisner (1997) shows that a reduced version of Optimality Theory is able to encode NP-hard problems (Garey & Johnson 1979); see Idsardi (2006) and Heinz *et al*. (2009) for further discussion of how to interpret these results and resulting research strategies. For the bottom right cell, there are no algorithms that can identify the class of regular relations in the limit from positive examples (Gold 1967). If phonological maps belong to this class, as is widely believed, then choices include identifying learnable subclasses of maps (Chandlee 2014) or rejecting aspects of the learning paradigm, such as the positive-only data aspect.

To summarise, all of the results mentioned above are important, but they do not come labelled with instructions for what to do about them.

5 General comments

All of the approaches covered in this special issue (data analysis, simulation, proofs) can make valuable contributions to the study of phonology. Obviously, researchers have their own favourite ways of working, and can disagree about the relative merits of each approach. Stabler (2014: 24), for example, is critical of simulation-based research:

> As in computing quite generally, running programs on particular examples to see what they do is usually much less valuable than carefully

considering what needs to be achieved and what kinds of computations could achieve that.

On the other hand, Niyogi (2006: 37–39) offers a more balanced perspective:

> Another aspect of the book is its focus on mathematical models where the relationship between various objects may be formally (provably) studied. A complementary approach is to consider the larger class of computational models where one resorts to simulations. Mathematical models with their equations and proofs, and computational models with their equations and simulations provide different and important windows of insight into the phenomena at hand. In the first, one constructs idealized and simplified models but one can now reason precisely about the behavior of such models and therefore be very sure of one's conclusions. In the second, one constructs more realistic models but because of the complexity, one will need to resort to heuristic arguments and simulations. In summary, for mathematical models the assumptions are more questionable but the conclusions are more reliable – for computational models, the assumptions are more believable but the conclusions more suspect.

While we are certain that modelling approaches will yield important insights, our own sympathies lie more with the Stablerian viewpoint, which could be summed up with the motto 'models are for now; proofs are forever'.

REFERENCES

Beesley, Kenneth R. & Lauri Karttunen (2003). *Finite state morphology*. Stanford: CSLI.

Chandlee, Jane (2014). *Strictly local phonological processes*. PhD dissertation, University of Delaware.

Daland, Robert (2014). What is computational phonology? *Loquens* **1**. doi.org/10.3989/loquens.2014.004.

Eisner, Jason (1997). Efficient generation in primitive Optimality Theory. In *Proceedings of the 35th Annual Meeting of the ACL and 8th Conference of the European Chapter of the Association for Computational Linguistics*. Morristown, NJ: Association for Computational Linguistics. 313–320.

Gallagher, Gillian, Peter Graff, Shigeto Kawahara & Michael Kenstowicz (eds.) (2012). *Phonological similarity*. Special issue. *Lingua* **122**. 107–176.

Garey, Michael R. & David S. Johnson (1979). *Computers and intractability: a guide to the theory of NP-completeness*. New York: Freeman.

Gold, E. M. (1967). Language identification in the limit. *Information and Control* **10**. 447–474.

Goldwater, Sharon & Mark Johnson (2003). Learning OT constraint rankings using a Maximum Entropy model. In Jennifer Spenader, Anders Eriksson & Östen Dahl (eds.) *Proceedings of the Stockholm Workshop on Variation within Optimality Theory*. Stockholm: Stockholm University. 111–120.

Hayes, Bruce & Colin Wilson (2008). A maximum entropy model of phonotactics and phonotactic learning. *LI* **39**. 379–440.

Heinz, Jeffrey (to appear). The computational nature of phonological generalizations. In Larry M. Hyman & Frans Plank (eds.) *Phonological typology*. Berlin & New York: Mouton de Gruyter.
Heinz, Jeffrey, Gregory M. Kobele & Jason Riggle (2009). Evaluating the complexity of Optimality Theory. *LI* **40**. 277–288.
Hulden, Mans (2009). Foma: a finite-state compiler and library. In *Proceedings of the 12th Conference of the European Chapter of the Association for Computational Linguistics: demonstrations session*. Association for Computational Linguistics. 29–32.
Hyman, Larry M. (1975). *Phonology: theory and analysis*. New York: Holt, Rinehart & Winston.
Idsardi, William J. (2006). A simple proof that Optimality Theory is computationally intractable. *LI* **37**. 271–275.
Jardine, Adam (2016). *Locality and non-linear representations in tonal phonology*. PhD thesis, University of Delaware.
Kaplan, Ronald & Martin Kay (1994). Regular models of phonological rule systems. *Computational Linguistics* **20**. 331–378.
Lau, Jey Han, Alexander Clark & Shalom Lappin (2016). Grammaticality, acceptability, and probability: a probabilistic view of linguistic knowledge. *Cognitive Science*. doi.org/10.1111/cogs.12414.
Niyogi, Partha (2006). *The computational nature of language learning and evolution*. Cambridge, Mass.: MIT Press.
Rogers, James (1998). *A descriptive approach to language-theoretic complexity*. Stanford: CSLI.
Rogers, James & Geoffrey K. Pullum (2011). Aural pattern recognition experiments and the subregular hierarchy. *Journal of Logic, Language and Information* **20**. 329–342.
Savitch, Walter J. (1993). Why it might pay to assume that languages are infinite. *Annals of Mathematics and Artificial Intelligence* **8**. 17–25.
Stabler, Edward (2014). Towards a rationalist theory of language acquisition. *Journal of Machine Learning Research: Workshop and Conference Proceedings* **34**. 21–32.
Tesar, Bruce & Paul Smolensky (2000). *Learnability in Optimality Theory*. Cambridge, Mass.: MIT Press.