Computation also matters: a response to Pater (2018)*

Adam Jardine, Rutgers University

October 20, 2018

Abstract

This article responds to Pater (2018) by arguing for a view of phonology that captures the computational properties of phonological processes. Jardine (2016)'s statement that tone is formally more complex than segmental phonology is not a claim, as Pater (2018) characterises it, but an empirical observation. This article thus outlines how phonological theories can incorporate such observations and integrate them with considerations of phonological substance. The conclusion is that, while computational characterisations are not necessarily alternatives to Optimality Theory, it is extremely difficult to capture the computational nature of phonological processes in Optimality Theory, due to the expressive power of global optimisation.

1 Introduction

Both substance and computation are central to a formal theory of phonology. *The Sound Pattern of English* (SPE; Chomsky and Halle, 1968) and Optimality Theory (OT; Prince and Smolensky, 2004) are both formal systems in that they specify computations over representations. What makes them phonological systems is

*To appear in *Phonology*. Author contact: adam.jardine@rutgers.edu. This paper is indebted to invaluable discussion with Jeff Heinz, Bill Idsardi, and the helpful comments of three anonymous reviewers. All errors are my own.

that their representations—inputs, outputs, candidates, rules, constraints—refer to phonological "substance"; that is, representational primitives that are specific to phonology. However, because they are formal systems, they each make claims about the kinds of computations that are possible for the phonological module of the grammar. In SPE, this computation is handled by a rewrite rule formalism, and in OT the central computational mechanism is parallel optimisation through the generation and filtering of candidates.

Jardine (2016) advocates for directly studying the computational properties of phonological patterns, arguing that the space of possible computations for segmental phonology is strictly smaller than that of tonal phonology. Pater (2018) criticises this conclusion for making incorrect predictions about the typology of tone and for being unable to make statements about phonological substance. He instead argues that OT provides a superior answer to both issues. The goal of this paper is to rebut these arguments by clarifying the computational view of phonology and by critically examining the claim that OT is superior at capturing the typology and incorporating phonological substance.

First, Jardine (2016)'s findings are not a final theory of either tonal or segmental phonology, but instead point to important properties that should be maintained by theories of phonology. Thus, when Pater (2018) claims that Jardine's characterisations conflate high-tone spreading in Copperbelt Bemba (Bickmore and Kula, 2013, 2015) with truly unattested "sour grapes"-type spreading (Wilson, 2003, 2006), this misunderstands the nature of the argument. Whatever our theory of tone ultimately may be, the fact remains that tone commonly includes patterns that are computationally different from those in segmental phonology. A computational view of phonology encourages future phonological theories to build on this fact.

In contrast, the power of optimisation makes it difficult to reconcile OT grammars with the computational nature of phonological processes. The phonological literature has already noted that classic OT theories of spreading make incorrect typological predictions, to the extent that they have motivated significant deviations from the classic OT architecture (Wilson, 2003; McCarthy, 2010; Rose and Walker, 2011). This can be directly connected to the fact that OT does not constrain the computational properties of processes—an extreme example of "mass-metathesis" is given below. For this reason, it is far from clear that Pater (2018) has provided a superior characterisation of the typology of spreading. As Pater points out, computational theories and OT are not necessarily incompatible. However, more work needs to be done to show how OT may be reconciled with computational characterisations of the kind highlighted by Jardine (2018).

Finally, substantive considerations can indeed be incorporated into computational characterisations of phonology. Much like substantive statements are formalized in OT grammars through specifying the content of CON, we can place substantive restrictions on computational characterisations through statements about the substructures and representations to which formal language-theoretic grammars refer. As detailed below, this is the topic of ongoing research.

This paper is structured as follows. §2 briefly reviews the relevant facts and the computational properties Jardine (2016) uses to explain them. §3 shows how OT does not adhere to these properties. §4 explains how substance can be incorporated into computational theories of phonology, and §5 briefly discusses exceptions to the generalisations discussed by both Jardine and Pater. §6 concludes.

2 The computational nature of phonological processes

The result of Jardine (2016) rests on studying phonological processes as functions from an input string to an output string. This approach reveals that segmental processes are surprisingly limited in the information needed to compute them: there is a bound on the amount any target can "look ahead" in the direction of application of the process in order to know whether or not it will change. This is expressed formally by placing segmental phonology within classes of functions related to the *subsequential* class of functions, a strict subclass of the *regular* class of functions for which the output can be computed deterministically by processing the input string left to right.[1]

A summary of bounds on processes and how that reflects on their status as functions is given in Table 1. Processes with a bound on both on the information to the left and to the right are *input strictly local* (Chandlee, 2014; Chandlee and Heinz, 2018). This is the case for any process which can be written in the form $A \rightarrow B \ / \ C \_\_ D$, where C and D are of some fixed size. Processes with a bound only on the right context, like unbounded nasal spread in Johore Malay (Onn, 1980), are left-subsequential. Processes with a bound only on the left context, such as long-distance consonant assimilation in Inseño Chumash (Applegate, 1972; Hansson, 2010), are right-subsequential. To accomodate bidirectional processes (such as root-control harmony; Baković 2000), Heinz and Lai (2013) posit the *weakly deterministic* class, which extends the subsequential class to include functions consisting of the application of a single subsequential function both left-to-right and right-to-left.

Position of bound	Type of Function
L & R contexts	input strictly-local
R context	left-subsequential
L context	right-subsequential
R context (parsing L→R), L context (parsing R→L)	weakly deterministic
None	beyond weakly det.

Table 1: Summary of types of function based on the boundedness of the contexts of a target (L=left, R=right). A dashed line indicates a region of complexity strictly more expressive than the region above it. The region highlighted in gray is hypothesized (Heinz and Lai, 2013; Jardine, 2016) to be excluded in segmental phonology.

[1]While this means that subsequential functions can be described by deterministic finite-state transducers, the property of subsequentiality is one of the function itself and is defined independently of how a function might be represented. Briefly, a subsequential function is one in which all input strings fall into one of some finite number of groups sharing a set of possible outputs (Schützenberger, 1977).

Thus, subsequentiality captures fundamental properties of the application of phonological processes. This characterization has found substantial empirical support. Typological studies have shown that local processes (Chandlee and Heinz, 2018), vowel harmony (Heinz and Lai, 2013), consonant harmony (Luo, 2017), and long-distance dissimilation (Payne, 2017) are all at most weakly deterministic. Heinz and Lai (2013) give additional evidence for the weakly deterministic hypothesis in the absence of sour grapes-style spreading (Wilson, 2003, 2006). They argue that it is not weakly deterministic, as it requires unbounded lookahead both to the left and the right of any potential target.

Even putative examples of 'non-myopic' harmony, as in Pater (2018)'s example of Central Veneto, can be subsequential. In Central Veneto, a stressed /e/ or /o/ raises when preceding a high final vowel, causing an intervening vowel to raise, as in (1a) below. That this only applies to stressed mid vowels is shown in (1b). Potential targets are underlined, and high vowels are bolded (data from Walker, 2010).

(1) a. órden-o **ú**rd**i**n-**i** 'order (1SG/2SG)'
    b. ázen-o ázen-**i** 'donkey (masc. SG/PL)'

This is considered non-myopic because because the intervening vowel must 'look ahead' to the preceding stressed vowel to see whether it should raise. However, stress in Central Veneto occurs on either the penult or antepenult (Walker, 2010). Thus, for a stressed vowel to find a high trigger for raising, it only needs to look ahead two vowels, and any intervening vowel only needs to look one vowel to either side to determine that it is in the right environment for raising. Thus, Central Veneto is also subsequential (in fact, it is input strictly local).

Jardine (2016) then gives evidence that non-weakly deterministic processes are commonly attested in tone, and thus posits that the weakly deterministic bound must crucially apply only to segmental processes. He cites unbounded tone plateauing (Kisseberth and Odden, 2003; Hyman, 2011) and unbounded spreading in Copperbelt Bemba (Bickmore and Kula, 2013, 2015) as examples of processes that require unbounded lookahead on *both* the left *and* the right of any given target. Importantly, while Pater (2018) cites this as a "claim" (abstract; p. 151) that tone is formally more complex than segmental phonology, it is actually a *fact*. Segmental phonology overwhelmingly is weakly deterministic, while unbounded tone plateauing and unbounded spreading in Copperbelt Bemba are outside of this space. As these are properties of the functions themselves, this fact holds independently of how they might be represented with a particular grammatical formalism.

Pater (2018) notes a difference between 'true' sour grapes and Copperbelt Bemba, and argues that the former is not attested in tone. If this is true, then our ultimate theory of phonology should exclude true sour grapes, even in tone. The right theory may provide a substantive answer, as Pater suggests, or it may provide a computational one. Regardless, it does not change the fact that tone commonly includes non-weakly deterministic processes. Thus, Pater's criticism of Jardine (2016) as not distinguishing between different non-weakly deterministic processes is misplaced.

3 Explanation and Optimality Theory

Pater (2018) argues for OT grammars because they can state substantive considerations, and claims to explain in OT why Copperbelt Bemba is attested in tone but "true" sour grapes spreading is not. However, without a theory of CON or an analysis of its factorial typology, such stipulations are only descriptive restatements of the empirical facts in terms of the framework, not a predictive theory of phonology.

Furthermore, the power of optimisation makes it difficult to reconcile OT grammars with the restrictive nature of phonological processes. Pater (2018) states that "most OT theories of spreading do not produce sour grapes" (p. 154), however the pathological predictions of parallel OT with regards to unbounded spreading are well-documented (Wilson, 2003, 2006; McCarthy, 2010; Rose and Walker, 2011). Thus, Pater's claim that he has demonstrated an OT system that "can generate the desired typology" (p. 155) is not sufficiently supported.

The reason for these issues is clear when we examine the expressive power of optimisation. Even with simple constraints, OT grammars can compute non-subsequential, fully regular functions such as sour grapes (Heinz and Lai, 2013) as well as non-regular functions such as majority rules (Gerdemann and Hulden, 2012; Heinz and Lai, 2013) and alphabetical sorting (Lamont, 2018).

To give an example, sorting can be produced by ALIGN-based spreading constraints. Take rightward spreading of a [nasal] feature, as discussed in §2. As pointed out by McCarthy (2010), ranking an ALIGN-R([nasal],word) constraint over LINEARITY predicts metathesis of a blocker and undergoer in order to more optimally align the [nasal] feature. In fact, the problem is worse than McCarthy illustrates: given a string of undergoers and blockers, *all* blockers are sent to the right edge of the word in a 'mass-metathesis' so that the winning candidate sorts the form into a nasal portion followed by a non-nasal portion.

(2)

/mawararara/	*NASLIQ	ALIGN-R([nas])	LIN	ID([nas])
☞ mãw̃ããããrrr	0	3	6	6
mãw̃ããrrara	0	5	1	4
mãw̃ãrarara	0	6	0	3
mãw̃ãr̃ãr̃ãr̃ã	3	0	0	9

The winning candidate in (2) is [mãw̃ããããrrr], in which LINEARITY has been massively violated so that all of the undergoers are ordered before blockers. This is not even a regular function, because the amount of memory required to compute the output grows with the number of blockers in the input that must be deposited at the end of the word. Thus, while there are OT theories of spreading that do not produce sour grapes, they can still miss the generalisation that segmental spreading is weakly deterministic (and even that it is regular).

Issues like sour grapes and mass-metathesis arise because optimisation allows direct comparison of candidates with local changes to candidates with non-local changes. This 'global' evaluation of optimisation is, as these examples show, computationally very powerful. It may be possible to rein in the power of OT grammars by carefully choosing the right constraints. We could reject AGREE, as suggested by Pater (2018). We would also have to reinterpret LINEARITY as an inviolable part of GEN, as the violability of LINEARITY is partially the culprit for the pathology in (2). However, all such modifications 'conspire' to capture a generalisation that is stated directly by the weakly deterministic hypothesis.

Furthermore, it does not appear to be the case that OT's problems with unbounded spreading can be solved by simply choosing the right constraint set. To capture the "myopic" nature of spreading, both Wilson (2003, 2006) and McCarthy (2010) appeal to significant changes to how OT evaluates candidates. In their review of harmony systems and their analyses, Rose and Walker (2011) tellingly note: "Proposals like those made by Wilson ... involve substantial departures from traditional constraint architecture in OT" (p. 268). The reason why Wilson (2003) proposes targeted constraints, which only assess a limited set of candidates, is because he finds that previous analyses of harmony in classical OT fail to capture the basic nature of segmental spreading. Weak determinism, in contrast, captures it directly, and provides a meta-theoretical bound for future theories to adhere to.

This is not to say that future work cannot develop versions of OT that are compatible with the weakly deterministic hypothesis. In an important sense, Pater (2018) is correct when he states that formal language theory is a "general tool for the formalisation and comparison of theories" and that computational theories of phonology are not necessarily an alternative to OT. Indeed, there is a body of work connecting OT to regular functions (Frank and Satta, 1998; Eisner, 1997; Riggle, 2004).

However, OT's difficulties capturing the computational nature of spreading highlight the need to further study the expressivity of OT grammars. Can we characterize a class of constraints that produce non-weakly deterministic (or non-regular) functions under optimization? Under what conditions, if any, can OT grammars be restricted to weakly deterministic functions? Research answering such questions could follow that of Tesar (2014), who proves the conditions necessary for constraint sets to produce typologies that are restricted to output-driven functions. Regardless, analytically studying the computational properties of OT grammars is critically important to evaluating it as a theory of phonology.

4 Substantive explanation and the weakly deterministic hypothesis

Finally, substantive explanation is not at all incompatible with computational theories. Pater (2018) assumes that any substance-based explanation cannot be reconciled with the weakly deterministic hypothesis: he states that FLT cannot provide "a means for the development of substantive theories of constraints" (p. 152). However, the hypothesis that segmental phonology is weakly deterministic is inherently a substantive one—it makes predictions about how phonological substance can interact in segmental versus other domains. There is no real difference between tone and segmental phonology in the substance Pater refers to—in terms of positional licensing, both tone and vowel features have been shown to refer to prominent positions. What the FLT analysis of their difference *explains* is that vowel features interact with that substance in a different way than tone.

More importantly, substance can be incorporated into FLT grammars through restrictions on the constraints grammars can choose from, and through the representations over which grammars operate. In other words, one can make substantive stipulations in FLT grammars much in the same way they are made in OT, or in any other theory of phonology.

This is most obviously true for phonotactic grammars based in FLT, which capture patterns through constraints that specifiy illicit substructures (Heinz, 2010; McMullin and Hansson, 2016; Jardine and Heinz, 2016), just as markedness constraints in OT (de Lacy, 2011). The only difference is that the constraints in these FLT characterizations are interpreted as inviolable. Thus, just as substantive statements can be made in OT by stipulating the content of CON, we can similarly make substantive statements in FLT phonotactic grammars through stipulations on what constraints are available to grammars. This is also true for processes. For example, we can study the range of functions that perform repairs on particular marked structures (Chandlee et al., 2015). Here also we can make substantive statements by specifying which structures need to be repaired.

Furthermore, as with other kinds of phonological grammars, substance can also be incorporated into phonological theories based in FLT through representation. This is implicit in the alphabets of symbols chosen in FLT-based work—Heinz (2009) uses strings of syllables to characterize stress, Heinz and Lai (2013) assume a vowel tier, and Jardine (2016) uses strings of moras. FLT-based studies of non-string phonological representations are the subject of ongoing research; see Jardine (2017) for tonal autosegmental representations and Strother-Garcia (2017) for syllable representations. In particular, the reader is referred to Jardine (2018), who argues that formal grammars defined over autosegmental representations more naturally capture the typology of tone well-formedness patterns than established formal grammars over strings.

5 Exceptions to the weakly deterministic hypothesis

It should be noted that non-weakly deterministic segmental processes, although extremely rare, are attested. One, discussed in Jardine (2016), is an apparent unbounded plateauing vowel harmony process in Yaka (Hyman, 1998). Another, which has come to light since Jardine (2016) was published, is ATR harmony in Tutrugbu (McCollum and Essegbey, 2018), in which unbounded harmonizing of low vowels to a following [+ATR] vowel is blocked by a preceding [+high] initial prefix.

Such patterns are of interest because they challenge a categorical interpretation of the weakly deterministic hypothesis. Indeed, McCollum et al. (2017) interpret the facts of Tutrugbu as evidence for the weakly deterministic bound as a bias, and not a categorical constraint (see also Avcu, 2018). However, it is also the case that neither are evidence for Pater (2018)'s OT-based prosodic licensing explanation of non-myopic processes. Hyman's analysis for Yaka invokes a sour-grapes style analysis, and McCollum and Essegbey invoke a combination of spreading and correspondence constraints. Thus, both processes are just as unexpected for existing OT-based theories of spreading. The upshot is that both cases warrant further investigation, regardless of one's theoretical perspective.

6 Conclusion

Phonological processes share non-arbitrary structural properties. A theory of phonology should thus be able to state these properties directly. Segmental processes are, almost without exception, weakly deterministic. This generalisation escapes explanation in OT because, as illustrated in §3, global optimisation makes it difficult, if not impossible, to make any coherent statement about the structural properties of phonological processes in OT.

FLT characterisations of phonology nontrivially capture generalisations that are important to any theory of phonology. At the least, then, they can serve as constraints on the kinds of computations that phonological theories should allow. Thus, when we look to include substantive considerations, we should do in a way that is in harmony with computational characterisations. Given the challenges of reining in optimisation, it is not clear that this is possible in OT.

References

Applegate, Richard B. (1972). *Ineseño Chumash grammar*. Ph.D. dissertation, University of California, Berkeley.

Avcu, Ennis (2018). Experimental investigation of subregular hypothesis. In *WCCFL 35*, 77–86.

Baković, Eric (2000). *Harmony, dominance and control*. Ph.D. dissertation, Rutgers University.

Bickmore, Lee S. and Kula, Nancy C. (2013). Ternary spreading and the OCP in Copperbelt Bemba. *Studies in African Linguistics*, 42.

Bickmore, Lee S. and Kula, Nancy C. (2015). Phrasal phonology in Copperbelt Bemba. *Phonology*, 32. 147–176.

Chandlee, Jane (2014). *Strictly Local Phonological Processes*. Ph.D. dissertation, University of Delaware.

Chandlee, Jane and Heinz, Jeffrey (2018). Strictly locality and phonological maps. *LI*, 49. 23–60.

Chandlee, Jane, Jardine, Adam, and Heinz, Jeffrey (2015). Learning repairs for marked structures. In *Proceedings of the 2015 Annual Meeting on Phonology*. LSA.

Chomsky, Noam and Halle, Morris (1968). *The Sound Pattern of English*. Harper & Row.

de Lacy, Paul (2011). Markedness and faithfulness constraints. In Marc Van Oostendorp, Colin J. Ewen, Elizabeth Hume, and Keren Rice (eds.), *The Blackwell Companion to Phonology*. Blackwell.

Eisner, Jason (1997). Efficient generation in primitive Optimality Theory. In *Proceedings of the 35th Annual Meeting of the Association for Computational Linguistics (ACL)*, 313–320. Madrid.

Frank, Robert and Satta, Giorgio (1998). Optimality theory and the generative capacity complexity of constraint violability. *Computational linguistics*, 24. 277–299.

Gerdemann, Dale and Hulden, Mans (2012). Practical finite state optimality theory. In *Proceedings of the 10th International Workshop on FSMNLP*, 10–19. ACL.

Hansson, Gunnar Ólafur (2010). *Consonant Harmony: Long-distance Interaction in Phonology*. Berkeley: University of California Press.

Heinz, Jeffrey (2009). On the role of locality in learning stress patterns. *Phonology*, 26. 303–351.

Heinz, Jeffrey (2010). Learning long-distance phonotactics. *LI*, 41. 623–661.

Heinz, Jeffrey and Lai, Regine (2013). Vowel harmony and subsequentiality. In Andras Kornai and Marco Kuhlmann (eds.), *Proceedings of the 13th Meeting on Mathematics of Language*. Sofia, Bulgaria.

Hyman, Larry (1998). Positional prominence and the 'prosodic trough' in Yaka. *Phonology*, 15. 14–75.

Hyman, Larry (2011). Tone: Is it different? In John A. Goldsmith, Jason Riggle, and Alan C. L. Yu (eds.), *The Blackwell Handbook of Phonological Theory*, 197–238. Wiley-Blackwell.

Jardine, Adam (2016). Computationally, tone is different. *Phonology*, 33. 247–283.

Jardine, Adam (2017). The local nature of tone-association patterns. *Phonology*, 34. 385–405.

Jardine, Adam (2018). The expressivity of autosegmental grammars. *Journal of Logic, Language, and Information*, (Online first). 1–46. ISSN 1572-9583. doi:10.1007/s10849-018-9270-x.

Jardine, Adam and Heinz, Jeffrey (2016). Learning tier-based strictly 2-local languages. *Transactions of the Association for Computational Linguistics*, 4. 87–98. ISSN 2307-387X.

Kisseberth, Charles and Odden, David (2003). Tone. In Derek Nurse and Gérard Philippson (eds.), *The Bantu Languages*. New York: Routledge.

Lamont, Andrew (2018). Precedence is pathological: The problem of alphabetical sorting. Poster, WCCFL 36.

Luo, Huan (2017). Long-distance consonant agreement and subsequentiality. *Glossa*, 2:52.

McCarthy, John (2010). Autosegmental spreading in Optimality Theory. In John A. Goldsmith, Elizabeth Hume, and W. Leo Wezels (eds.), *Tones and Features: Phonetic and Phonological Perspectives*, 195–222. De Gruyer Mouton.

McCollum, Adam, c, Eric Bakoví Mai, Anna, and Meinhardt, Eric (2017). Conditional blocking in Tutrugbu requires non-determinism: implications for the subregular hypothesis. In *Paper presented at NELS 48*.

McCollum, Adam G. and Essegbey, James (2018). Unbounded harmony is not always myopic: Evidence from Tutrugbu. In Wm. G. Bennett, Lindsay Hracs, and Dennis Ryan Storoshenko (eds.), *WCCFL 35*, 251–258.

McMullin, Kevin and Hansson, Gunnar Ólafur (2016). Long-distance phonotactics as Tier-Based Strictly 2-Local languages. In *Proceedings of AMP 2015*.

Onn, Farid M. (1980). *Aspects of Malay Phonology and Morphology: A Generative Approach*. Kuala Lampur: Universiti Kebangsaan Malaysia.

Pater, Joe (2018). Substance matters: a reply to Jardine (2016). *Phonology*, 35:1. 151–156.

Payne, Amanda (2017). All dissimilation is computationally subsequential. *Language: Phonological Analysis*, 93:4. e353–371.

Prince, Alan and Smolensky, Paul (2004). *Optimality Theory: Constraint Interaction in Generative Grammar*. Blackwell Publishing.

Riggle, Jason (2004). *Generation, Recognition, and Learning in Finite State Optimality Theory*. Ph.D. dissertation, UCLA.

Rose, Sharon and Walker, Rachel (2011). Harmony systems. In John Goldsmith, Jason Riggle, and Alan Yu (eds.), *The Handbook of Phonological Theory*, 240–290. Wiley-Blackwell.

Schützenberger, Marcel Paul (1977). Sur une variante des fonctions séquentielles. *Theoretical Computer Science*, 4. 47–57.

Strother-Garcia, Kristina (2017). Imdlawn Tashlhiyt Berber syllabification is quantifier-free. In *Proceedings of the first annual meeting of the Society for Computation in Linguistics*, volume 1, 145–153.

Tesar, Bruce (2014). *Output-Driven Phonology: Theory and Learning*. Cambridge Studies in Linguistics.

Walker, Rachel (2010). Nonmyopic harmony and the nature of derivations. *LI*, 41. 169–179.

Wilson, Colin (2003). Analyzing unbounded spreading with constraints: marks, targets, and derivations. Unpublished ms.

Wilson, Colin (2006). Unbounded spreading is myopic. Paper presented at the workshop Current Perspectives on Phonology.