History of Phonology: Optimality Theory

Marc van Oostendorp

Almost immediately after the first manuscript (Prince and Smolensky 1993) started circulating, Optimality Theory (OT) arguably became the lingua franca of phonological theory, and to some extent it serves this function until today.[1] This chapter aims to describe how this could have happened, and in particular what the consequences of this were, for OT and for phonology: more and more power was added to the notation as virtually any idea that phonologists thought valuable has been expressed by it.

Although quite some work has been done on the theory itself as well in the past 25 years (see e.g. Brasoveanu and Prince 2011), and as a consequence we seem to have a more precise view of many of its formal properties, and although we will see below that several areas of research have achieved success by keeping to the initial restrictions, the simultaneous use of Optimality Theory as a lingua franca means that it is difficult to find any clear restriction on what constitutes a possible analysis in the body of OT literature that has accumulated over the past 25 years. Such may be the fate of any theory that becomes mainstream in our field, given the lack of consensus on many fundamental issues.

An additional complication is that OT was never a full-fledged theory of phonology to begin with. Already in its initial form, which was instantiated in Prince and Smolensky's manuscript, and which we will refer to as Classic OT in this paper, it did not put forward any inherent claim about representations and (therefore) of constraints. OT is primarily a theory of variation internal and external to languages, and most debates within the theory have centered on these topics. It has been the relative neutrality of OT with respect to many core issues of phonological theory that may have led to its acceptance as a lingua franca; it has been its focus on variation and computation which turned these into one of the best studied topics in the field in the past few decades.

In essence, OT claims that this computation consists of two functions, called Gen and Eval. Gen takes a phonological input (which is any phonological object) and converts it to a set of (infinitely many) candidate output forms. Eval takes this set and selects one candidate, the winner, which will be the output of the grammar, based on evaluation of a set of constraints. The set of inputs, the functions Gen and Eval, and the set of constraints are all assumed to be universal. The only language specific aspect of the grammar is a ranking of the universal constraints, which can conflict and are violable. Most of these assumptions within Classic OT have been questioned in the literature, however, as we will see.

The theory offers thus a view on computation - there is in principle only the universal derivation in terms of the general functions Gen and Eval - as well as a theory of grammatical variation, which has a very clear locus. It does not provide a view on other things, e.g. on syllable structure or the way in which phonological features interact. In their manuscript, Prince and Smolensky made certain choices in this respect, but they also made it clear that those choices were independent of their actual proposals.

An important aspect of the theory is its notation. Candidate analyses and the rele-

1Thanks are due to Edoardo Cavirani, Björn Köhnlein and Tobias Scheer for useful comments.

vant constraints are presented in a table that is called a 'tableau', with asterisks marking violations of constraints by candidates, and pointing fingers indicating winning candidates:

(1)

	Constraint A	**Constraint B**	**Constraint C**
output a	*!		
output b		**!	
☞ output c		*	*

The constraints are ordered left to right in the order of their dominance: A occurs on the left of B because a violation of A is considered more important than a violation of B. In this example, output c 'wins' because it does not violate the highest ranked constraint A (like output a), and violates the constraint B only once (whereas the competing output b violates it twice). The exclamation marks indicate 'fatal' violation and the pointing finger points to the winner. The fields that are irrelevant to the final computation have been shaded. The notation has in some ways remained more stable than the actual theory (the shading for instance seems to have largely disappeared, probably because it is redundant with the exclamation marks). There have been proposals to improve it (e.g. in Prince 2002a), but these have not been generally accepted.

I believe that the fact that OT inherently does not provide a full theory of phonology cannot be held against it; the same is essentially true for most phonological 'theories'. For instance, Autosegmental Phonology (Chapter *) does not provide a theory of phonological computation – how inputs and outputs are related – and neither does Government Phonology (Chapter *). Certain frameworks seem to provide more of a complete picture – SPE (*The Sound Pattern of English,* Chomsky and Halle 1968) comes to mind – but even in those cases we can wonder whether different components of the analysis are inherently related, e.g. in the case of SPE whether the assumption of segments as feature bundles is necessarily connected to the idea that the derivations consist of extrinsically ordered rules. It has been the fate of phonological theory also outside of OT that basically every assumption has been under discussion at any point, and independent from all other points, without a lot of consensus on anything. This also implies in my view that one cannot dismiss OT simply for its lack of attention to representational detail – the individual analyst simply needs to complement it with a different theory, about which one can have by and large a separate discussion (see e.g. Blaho et al. 2008). Of course, OT can still be wrong on the things which it is a theory about.

I will argue in this chapter that the theoretical developments within OT have mostly gone in one direction: enlarging the descriptive power of the notational framework by loosening restrictions that were built into the original model. It is difficult to find successful proposals in the literature trying to make a *more restrictive* version of OT. This simply seems not to have been among the set of goals which the mainstream of researchers has set itself; as far as I can see all current practitioners of the framework have loosened some of the initial restrictions, and

in many cases quite radically so. In terms of the well-known terminology of Chomsky (1965), generally the focus of phonologists working on OT seems to have been on descriptive adequacy rather than on explanatory adequacy. Many OT papers start from the observation that some pattern in some language is difficult to describe with the existing tools and then proposes a new tool to facilitate that description. Very few papers show that some pattern is not found and therefore the theory needs further restrictions.
This research culture, which is markedly different from for instance the one we find in a contemporary (albeit less popular) theory like Government Phonology, might be both the result of OT's use as a theoretical lingua franca and a reason for that success: the theory in its most blown-up form (with add-ons for describing continuous variation, for rule ordering, etc.) offers a somewhat formalised notation system for describing phonological computation and variation that seems to offer no problems for most types of attested phenomena. For many phonologists, OT offers a formalism for writing theories rather than anything else.
The OT literature has been very broad in scope over the last 25 years; it is impossible to do justice to all details. To a large extent, the history of OT has been the history of theoretical phonology between 1993-2018, since even those theoreticians who did not subscribe to OT felt compelled to give their opinion about it. To just give one example, De Lacy (2007) presents itself as a 'Handbook of Phonology', but which only contains papers using OT. My goal in this Chapter is to concentrate on some of the tendencies I see, taking into account developments in different parts of the world.

1. Initial reception

Optimality was in the air, in the early 1990s. The field (or at least the generative subfield, but this included the majority of scholars at the time) had been moving away from the extrinsically ordered rules that SPE had proposed for a while. In the late 1970s and throughout the 1980s, most theories concentrated on representations, such as Autosegmental Phonology and Metrical Phonology (see Chapter *). This work implicitly or explicitly still assumed that rules were governing the relation between input and output, but according to a famous dictum by John McCarthy (1988) "if the representations are right, the rules will follow". The representations seemed to be sufficiently rich to make a complicated derivational machinery less relevant. An important demonstration of this could be found in the work on Prosodic Morphology by McCarthy and Prince and many others (see an overview in McCarthy and Prince 1996), showing how the representations independently established in Prosodic Phonology (Nespor and Vogel 1986) could also explain phenomena such as reduplication and infixation.
It may also have been relevant that generative syntax had moved away from a formalism that was very similar to SPE – that of *Aspects of the Theory of Syntax,* Chomsky (1965) – already during the 1980's which shifted the focus from rules to a modular system of constraints on representations such as Binding Theory and Case Theory. This also included an attractive and simple theory of language variation: parameter theory, that was also successfully applied to (metrical) phonology. It was not entirely clear how other aspects of the new syntactic framework could be applied to the domain of phonology, but there were several attempts towards a theory in which constraints on representations played a

more important role. (The idea of 'constraints' was known in generative phonology at least since Halle 1959's use of 'well-formedness conditions', see Cavirani 2015 for an overview)
One example of this was the work by Burzio (1994) on English stress, arguing for an analysis in terms of ranked and violable constraints – key properties of OT, which he seems to have found more or less independently. Kager (1995) in a review claims as a 'reservation' that Burzio should have mentioned Prince and Smolensky (1993) more (there is one reference to the manuscript in Burzio's book), but it is not entirely clear to me how much aware the author was of this work while writing his own book. Burzio in later work more explicitly adopted OT as a framework.
Another influential proposal is the so-called Theory of Constraints and Repair Strategies (TCRS) by Carole Paradis (see Chapter *), which, as the name suggests, combines the two main competing candidates for phonological computation. There is a set of constraints, which comes in two types: principles and parameters (TCRS thus is a P&P framework). In addition, constraints can work in two ways, on a language specific-basis: they can block the application of certain (also language-specific) rules, or they can trigger such rules when they are violated (see also McCarthy and Alves 2017).
The third proposal we can mention is Grounded Phonology of Archangeli and Pulleyblank (1994), in which rules are universal in the sense that they are 'phonetically grounded', but a language can make a parametric choice. Grounded Phonology touched on a topic of 'naturalness' that was of great concern to phonologists since Chapter 9 of SPE: how can we make sure that the 'naturalness' of phonological processes can be expressed?
The three proposals mentioned were definitely not the only ones; Odden (2011) gives an overview. Together they point to the kinds of concerns that seem to have played a role in the acceptance of Optimality Theory, such as the role of constraints and naturalness in phonology. OT seemed to combine those different aspects in a relatively simple theory that could easily be applied to a lot of phenomena.
From the point of view of the sociology of the field, it may be relevant that Prince and Smolensky mentioned almost all of these works and their authors as precursors to their own work. (As a personal aside, I should mention that in my own experience in the Netherlands, OT was initially mostly used by PhD students and postdocs. I was a PhD student myself and one of the attractions of the theory was that it was promising and new. Alan Prince visited the Netherlands twice and seemed to talk a lot to relatively junior linguists. OT for a short while was something mostly of the younger scholars.)
It may also have been relevant that almost at the same time, another manuscript on Optimality Theory started circulating that was titled Prosodic Morphology I, and written by Alan Prince and John McCarthy (1993), adding to the impression that Optimality Theory was a development of Prosodic Morphology, which was something that needed to be studied.
A final factor that we may mention is that the community immediately started using the internet: the so-called Rutgers Optimality Archive is an online archive which accepted manuscripts by all scholars contributing to the theory. It was the first such archive in (theoretical) linguistics and it made manuscripts very accessible, especially to junior researchers who until the mid 1990s had to rely on e.g.

their supervisors to get access to unpublished work. It continues to work, and contains at this moment (January 2018) around 1335 manuscripts.
In any case, OT took the field by storm. Within a few years, almost every handout at almost every conference was filled with the tableaux that are the graphical trademark of the theory, a phenomenon that can probably only be compared to the reception of SPE approximately 25 years earlier. Already six years after circulation of the first manuscript, and before its publication in 2004, there was a textbook in the prestigious 'red' Cambridge University Press series of textbooks (Kager 1999). Not everybody adopted the framework and there was opposition to it from a number of different points of view, but it is clear that in the 1990s and 2000s it was hard to be a practicing phonologist without at least a passive knowledge of Optimality Theory.

2. The opacity debate

The main discussions on the validity of Optimality Theory centered around the question whether it was able to deal with so-called opacity phenomena. The term is coined by Kiparsky (1971), who defined it as follows (very few definitions will have been cited as often in the OT literature):

(2) A rule A → B / C __ D is opaque to the extent that there are surface representations of the form:
  a. A in environment C __ D, or
  b. B in an environment other than C __ D

In Kiparsky's definition, opacity is a property of rules. One could therefore maintain that it is irrelevant for a theory which, like (Classic) OT, does not recognise the rule as a tool of analysis: if there are no rules, there can not be opaque rules. However, the term is nowadays used also for sets of phonological data – those that would typically require opaque rules in a rule-based analysis. As far as I can tell, this has been made explicit for the first time by Chomsky (1995) (in a manuscript which circulated already a while before that), referring to data in Tiberian Hebrew that had been the topic also of Chomsky's MA thesis (published as Chomsky 1979). These data involve spirantisation of intervocalic *k*. (3a) shows that the root for 'king' contains such a k; (3b) shows how it is spirantised intervocalically:

(3) a. *malki* 'king', *malka* 'queen'
  b. *mlaxim* 'kings'

There are however some forms in which we find a spirantised *x,* in spite of the fact that the consonant is not intervocalic on the surface, for instance in the so-called construct state:

(4) *malxey* 'kings of'

In terms of Kiparsky (1971), the intervocalic spirantisation rule is opaque; we find a 'B in an environment other than' the one that usually triggers spirantisation: there is a spirantised *x* in a non-intervocalic context. A rule-based analysis of this posits that there has been an (epenthetic) vowel preceding the consonant

in the construct state at some stage of the derivation, which is deleted after spirantisation took place. The rule thus applies normally, be it that its context has been erased by a later rule. That is what makes it 'opaque'. There is no way to mimic this analysis in Classic OT, since the epenthetic vowel can only be there at an intermediate stage of the derivation, which the theory does not have. (The issue was first brought up in a footnote in Chomsky (1995), and elaborated in Idsardi 1997).

The significance of this these phenomena is obviously that they can be accounted for in terms of 'opaque rules' (which in turn means: by rule ordering) in, for instance, *SPE*-inspired formalisms, but not by Classic OT. One can in principle respond to this state of affairs in a variety of ways. One of them is to deny that the phenomena belong to the domain of phonological theory. One could argue that instead they are the responsibility of historical phonology, of the morphology or of the lexicon (with respect to the Hebrew facts, this approach was for instance suggested by Green 2004).

Yet another way of approaching these phenomena is acknowledging that such data are to be accounted for by phonological theory, but denying that the general scheme in (2) is real: there is no generalisation over sets of data that warrants opaque rules. Maybe we have to account for the Hebrew data, but we do not have to account for them in the same way as we do for other phenomena that are also 'opaque'. This line has been pursued by a number of (mostly European) authors, who argue that one should try to find the solution to specific opacity problems in more sophisticated phonological representations, which for instance have elements in them which are not pronounced but can still be visible to the constraints (Goldrick 2001, Revithiadou 2007, Trommer and Zimmermann 2014, Zimmermann 2017). The kinds of representations that are needed may vary from one phenomenon to the next. Although the Hebrew example has not been discussed in this way, as far as I know, an analysis along these lines would hold that the surface representation of *malxeɣ* contains an inaudible vowel preceding the *x*, so that a constraint *VkV ('no intervocalic *k*') would be 'seeing' a vowel preceding it and still be violated if the underlying /k/ does not spirantize. Such proposals are more restricted in scope – probably not all types of 'opacity' according to the rule-based scheme can be analysed in this way. It is, as far as I can see, still an open empirical question whether this limited way of accommodating 'opacity' is warranted and it is still an open theoretical question whether that would be preferable to having a uniform mechanism covering all instances of Kiparsky's abstract scheme in (2).

In any case, such strategies have not become part of the 'mainstream' in Optimality Theory. Instead of this, a variety of approaches has been attempted to extend Classic OT in such a way that it would have the full power of SPE's extrinsic rule ordering. This is clearest in a string of influential works by John McCarthy, who proposed 'Sympathy Theory' (McCarthy 1999), 'Comparative Markedness' (McCarthy 2003), 'OT with Candidate Chains' (McCarthy 2007) and finally 'Harmonic Serialism' (McCarthy 2010a, see also section 4), which were extensions of Classic OT that became ever more powerful (with the possible exception of the step from Candidate Chains to Harmonic Serialism). In particular, Candidate Chain Theory has a way to implement extrinsic rule ordering, and it seems that one could literally translate any analysis in terms of ordered rules in terms of that theory. The two theories thus have become equivalent in their power.

3. Theories of faithfulness: Containment and Correspondence

Since Optimality Theory is, among other things, a theory of input-output mapping, it needs a way to implement 'economy': forms should only change from input to output when this is necessary and random changes should not be allowed. In rule-based formalisms one does not have to worry about this, since rules are the only operators causing change. If no rule is applicable to a form, nothing will happen. Any theory which does not have explicit rules, however, needs to make sure that changes apply only when necessary.

An original contribution to this discussion by Prince and Smolensky (1993) was that these principles of economy – which they called 'faithfulness' – are to be seen as violable constraints, and even as constraints on surface structures: both epenthetic segments (which are not present in the input but are added in the output) and deleted material (present in the input, but not pronounced) are supposed to be present in the phonological output structure, and marked in a special way. 'Faithfulness' constraints, then, were constraints against those special markings. By inspection of an output candidate form, one could thus see immediately whether it violated 'faithfulness', just like one could see whether it violated 'markedness', the term for the 'other' kind of constraints. Classic OT for this reason is a purely monostratal theory: in principle everything that is relevant can be read off one representational level only.

The view on faithfulness in Classic OT, called *Containment* (because input representations were fully contained in the output) was the first component of the theory to be almost universally abandoned within the OT community. As far as I can see this happened on the basis of one paper: Prince and McCarthy (1995). This paper introduced a different view of faithfulness that became immediately the new standard and was called *Correspondence.* In this view, the Eval function does not just take the output structure but also the input structure into account when evaluating the optimal candidate. Correspondence relations are established between elements of the two. An epenthetic segment thus is a segment in the output that has no correspondent in the input, whereas a deleted segment is a segment in the input that has no correspondent in the output. Faithfulness constraints require segments in both representations to have equivalents in the other representation. Correspondence thus is no longer monostratal as it crucially refers to two levels of representation. In return, the output representation obviously no longer has to contain special markings for epenthetic and deleted material.

The main initial argument for adopting such a view was that it allowed the analyst to generalize over input-output relations and the relations between the 'base' and the 'reduplicant' in reduplicated structures. Also in the latter case, McCarthy and Prince (1995) showed that segments can sometimes be 'inserted' or 'deleted', but only when some other (markedness) constraint makes this necessary. Containment-based analyses cannot represent this similarity between what happens in a reduplicated form and what happens in input-output relations in a straightforward way.

Since it adds a complete extra representation and a set of relations between elements of two representations, Correspondence Theory is a stronger theory, which can also account for a lot of phenomena that seem more difficult to analyze

under Containment in Classic OT. The richness of the theory has been further explored in later literature, where e.g. Correspondence was also allowed with other candidates (McCarthy's 1999 'sympathy theory', for opacity), or between segments in the same word in order to account for vowel harmony (Krämer 2008) as well as for (long-distance) assimilation (Walker 2000 and much following work). As far as I know, there is no other phonological theory claiming that all those relations are similar in kind. On the other hand, the work on long-distance assimilation has shown that the mechanism is more powerful than autosegmental phonology or other representational mechanisms.

4. Theories of derivations: Stratal OT, Harmonic Serialism

Classic OT championed a very shallow theory of the phonological derivation that consisted basically of only one step, from input to (candidate) output, and in which (because of Containment) this step did not even very matter much, since epenthesis and deletion were both representationally marked. We have already seen in section 2 that this led several researchers to worry because in this way it seemed impossible to deal with opacity phenomena that are previously covered in other theories by rule ordering.

Opacity is however not the only phenomenon that seems to require derivational ordering of some type. Almost from the beginning, some scholars proposed that the stratal organisation of 'stem-level', 'word-level' and 'phrase-level' phonologies needed to be preserved in order to account both for the differences between phenomena occurring in each of these levels, as well as for the fact that these phenomena seem to be inherently ordered, as is known from Lexical Phonology (Kiparsky 1982). The possibility is actually mentioned in Prince and Smolensky (1993:143). The proposal was always the same at some level of abstraction – to string together a set of OT systems, each devoted to a different level –, although the proposals initially came with a set of different names, such as Derivational OT (Rubach 2000; see Booij 2018 for some of the history) and Stratal OT (Bermúdez-Otero 2018). My impression is that the latter term in the end has stuck, mostly through the work of Bermúdez-Otero.

Stratal OT provides a morphologically motivated way of giving a derivational order between certain processes, but it claims that there can be no ordering of phonological processes within a stratum, just like in Classic OT, of which it is therefore a minimal departure.

Some researchers took a different path; in particular at the University of Massachusetts at Amherst this led to a lot of interest in a model (mentioned in a footnote in Prince and Smolensky 1993:56) that is called Harmonic Serialism (McCarthy 2010a). This model has one difference with Classic OT or Stratal OT: instead of applying many changes at the same time, only one change (e.g. deletion of one feature, building one syllable, etc.) is applied to give a candidate output, and the most optimal one is chosen, after which again one change is applied. This gives an ordered derivation without rules, since every change is chosen only because it satisfies the constraint ranking best.

It has been pointed out (McCarthy 2010b, Torres-Tamarit 2015) that the fact that Harmonic Serialism depends on the notion of 'one change' turned researchers back to issues of representation, and led them to study more seriously the working of the Gen function, since we can only count changes if we have a view of

what the representations look like. Harmonic Serialism has been the object of theoretical study for approximately the last 10 years, mostly by John McCarthy and his students. I see it as one of the few examples where a more restrictive model has won over a model with more power.
In a recent interview (McCarthy and Alves 2017), McCarthy points out that Harmonic Serialism also has other advantages over Classic OT. For instance it can solve certain instances of the 'too many solutions problem': in principle there sometimes can be many ways of satisfying some markedness constraint, but of these languages sometimes only use one, or a limited number. Harmonic Serialism can solve this problem by showing that some of the alternative 'solutions' not chosen by natural languages, would require wild, non-optimising jumps in the derivation. In any case it seems to me that Harmonic Serialism is one of the few examples where a theoretical change in the architecture of the theory was chosen because it restricted the empirical power of the theory rather than (only) enhance it.

5. Theories of constraint violation: Classic OT vs. Harmonic Grammar

Another aspect of OT that has been under discussion is the way in which constraints can be compared to each other. In any version of OT, if a constraint A dominates another constraint B, a form that violates A will always lose from a form that does not, no matter how many times the other form violates constraint B. In other words, if A dominates B in OT, violation of A is infinitely worse than violation of B.
Harmonic Grammar (not to be confused with Harmonic Serialism) proposes a variety of this in which every constraint gets a weight assigned to it that is some finite number. For instance, constraint A could get weight 2 and constraint B weight 1. In that case, if a candidate violates A it loses from any candidate that satisfies A and violates B one or two times, but it wins as soon as the other candidate violates B more than three times.
Harmonic Grammar is arguably older than Classic OT; it was proposed in Legendre, Miyamata and Smolensky (1990), and Prince and Smolensky apparently built their theory on this conceptual basis (see Smolensky 2006 for an overview). From the point of view of generative grammar, the move from Optimality Theory to Harmonic Grammar is however a big one, as concrete numbers have typically been avoided within that family of theories (Halle 1957). OT is actually mathematically a special case of Harmonic Grammar: any OT analysis can be straightforwardly translated into HG, provided the weights of the constraints are set at an appropriate distance, but the opposite is not necessarily true. This means that Harmonic Grammar offers fewer restrictions than OT on the number of possible analyses. The move from Harmonic Grammar to Classic OT thus seems a logical one from a classical generative point of view, aiming a restrictive models to maximise explanatory adequacy. The move in the opposite direction can only be justified on empirical grounds.
On the other hand, it has been pointed out (Pater 2016) that the stronger power of the constraint interaction mechanism means that individual constraints can be simpler in HG than in OT. For instance, Japanese devoices geminates in loanwords *iff* there is another voiced obstruent in the same (loan)word; if a word contains two non-geminate voiced obstruents neither of these is devoiced. In Harmonic

Grammar, this can mean that the constraint against voiced geminates and the constraint against two voiced obstruents in a word (OCP) each are dominated by a 'faithfulness' constraint in favour of preserving underlying voice, but they can 'gang up' to make that constraint less relevant. In OT, one would need some extra, and complicated, constraint (for instance formed by 'constraint conjunction, Crowhurst 2011) in the hierarchy to get the same effect. More generally, because constraints can interact in subtler ways in Harmonic Grammar, each of the constraints themselves can be very simple. It is difficult to see how to objectively make the tradeoff between simple constraints in a powerful system of interaction versus powerful constraints in simple interaction; but it should be noted that although Harmonic Grammar allows for simpler constraints in *principle*, there is nothing in the theory that forces the analyst to only use those simpler constraints, and I am not aware of attempts to precisely delineate which constraint types should not be allowed in Harmonic Grammar.

This touches on a more general issue with OT and OT-related models: that there is nothing like a generally accepted theory of what constitutes a possible constraint. In principle, any declarative statement can be labeled with a constraint name, and inserted into a tableau. In actual practice, people will tend to adhere as much as possible to constraints that have been proposed already in the literature, but there is surprisingly little attention to what would constrain the constraint set. The closest we have are some taxonomies of constraints in terms of 'faithfulness constraints', 'markedness constraints', 'alignment constraints', etc. But also this has hardly ever been a very lively field of research. (Ashley et al. 2010 give a good overview and taxonomy of constraints in the literature for the first 15 years; I have found no data about the later period.)

As I mentioned in the introduction to this Chapter, this is not an argument against OT per se, which is simply not a theory about the substance of constraints. But since the validity of individual analyses depends to a large extent on the constraints which are used, it is surprising that the community as a whole has not tried to establish what kinds of constraints are supposed to be in the constraint set. It is, at least in my view, an argument against individual analyses that they do not provide a theory about possible constraints.

Even the issue of the origin of the constraints in the constraint set is hardly discussed. Prince and Smolensky (1993) assume that the set of such constraints, Con, is part of Universal Grammar and thus provided by the innate language faculty. This is part of Classic OT's strong hypothesis that the variable factor between languages is the ranking of constraints. Although the issue is rarely discussed explicitly in the literature, one gets the impression that the idea of a rich UG has become much less popular and that many scholars nowadays assume that constraints are not necessarily innate, but somehow acquired by the learner, in a way that makes them grounded in the phonetics. This should make the demand to set up a syntax of possible constraints even more pressing – if the learner needs to automatically derive the constraints from phonetic substance, she needs to know how to do that –, but the issue has not come up for reasons which I find difficult to explain.

6. Theories of variation: from unranked constraints to stochastic grammar

Next to a theory of computation, OT is primarily a theory of language variation,

making a precise claim about how languages can be different, but also how there can be systematic variation within a language: how dialects or sociolects can differ and even how individual speakers can show internal variation, sometimes using one form and at other times using another form.
It should be noted that OT inherently has a more precise view on this than SPE phonology, in which languages could differ in having different rules, or small variations of rules, or different orderings of rules. How such variability should be accounted for within a speech community was studied quite extensively by William Labov and his students (Cedergren and Sankoff 1974), who introduced the variable rule format. This work hardly got integrated in the theoretical literature, however.
This has been very different in OT, where the study of variation, also at the individual level, seems to have become an integral part of the theoretical enterprise. There has been quite some work trying to formalise the grammar so that it could account precisely for the precise distribution of forms, and this in a variety of ways.
One line of work suggests that variable grammars contain constraints that are unranked. Arto Anttila has probably been the most explicit proponent of this idea (Anttila 2007 gives an overview), pointing out that one can derive predictions about the statistics in a corpus from this lack of ranking. Suppose we have three constraints A, B, C that are not ranked with respect to each other. We assume that this means that every ranking of these constraints (A≫B≫C, A≫C≫B, B≫A≫C, B≫C≫A, C≫A≫B, C≫B≫A, where A≫B means that A dominates B in OT notation) is equally likely. Now suppose that A≫B≫C, and A≫B≫C give one surface structure and the other rankings another one for some input. The prediction than is that the second structure is two times more likely than the first one. Anttila showed how this prediction is borne out in certain cases of 'free variation' in Finnish.
While Anttila's analysis seems compatible with Classic OT, some authors have taken a more radical step. Most successful among these seems to have been Stochastic OT (Boersma and Hayes 2001 and much subsequent work), a version of the theory in which the ranking of constraints was subjected to stochastic variation: for any pair of adjacent constraints A and B, it is specified how likely it is that A dominates B. Every time a form is evaluated, we can thus have a different ranking, according to this initial distribution.
Because it introduces real numbers in the description of phenomena, Stochastic OT can be seen as an OT version of the variable rule format. One difference is that it has become more integrated into the theoretical literature; papers appeared in 'hardcore' theoretical journals such as *Linguistic Inquiry*.
My impression is that another difference is that Stochastic OT has been less enthusiastically used in sociolinguistic circles, however, possibly because of a move away from formal grammar in the sociolinguistic mainstream (Nagy 2013). Another possible reason is that the formal analyses are often based on the assumption that the variation is 'free', i.e. not subjected to extra-grammatical (in particular sociolinguistic) constraints. It is not easy to make this connection; one would need to state that certain extralinguistic circumstances give extra weight to certain constraints, but as far as I am aware this avenue has not been explored in much detail (see however Coetzee 2016). The notion that variation is ever truly 'free' (i.e. random) is alien to most people interested in variation; it is also diffi-

cult to come up with examples where the choice between two variants does not seem to correlate to any matter of style, or social class, or any other factors. But if this is the case, it may in the end be misleading to ascribe such variation to 'stochastic' factors.
The development from Classic OT to stochastic OT led some scholars to set an extra step, and move on to MaxEnt grammars (e.g. Goldwater and Johnson 2003, Hayes and Wilson 2008), a system that is even more flexible (Magri and Anttila 2018) but for which the mathematical (computational) properties are quite well known as Maximum Entropy models have been studied widely also outside of linguistics (e.g. in bioinformatics). Other developments come from the fact that obviously stochastic mechanisms can also be implemented in sister formalisms of Classic OT, such as Harmonic Grammar. We then get a rather complicated model in which constraints have a ranking value that consists of some real number representing its average or median position in the hierarchy plus or minus some stochastically determined second value.

7. Theories of exceptions: indexed constraints, cophonologies

There is yet another type of variation that has also been studied within OT: not between speakers, or of speakers in different situations, but between different words. The fact that some words are 'exceptions' to otherwise general rules has been well-known in phonology, as well as the fact that certain generalisations only apply to certain subclasses of words, and this particular field of study is therefore not new to OT. There is, however, a problem rather specific to the OT conception of grammar with these data: since all variation is supposedly formalised as a different ranking of constraints, we need to assume that different rankings apply to different words.
This idea has been worked out in two different ways. The first, called 'cophonologies' (Ogun 1996, Inkelas and Zoll 2005), takes this idea literally and indeed assumes that there are (partially) different rankings available within a language. Individual lexical items can get a diacritic assigning them to one such 'cophonology' or another. The other idea, usually referred to as 'indexed constraints', posits that there is only one grammar, but that constraints can be duplicated, again with a diacritic mark. In other words, suppose one set of words need the ranking A≫B, and another set the ranking B≫A, the two theories both assume that the first set of language has, say, index *i*, and the other set, say, index *j,* and they give the following partial grammars:

(5) *cophonologies*
i: A≫B
j: B≫A

(6) *indexed constraints*
Ai≫B≫Aj

In the indexed constraints account, Ai is only applicable to words of type i and Aj only to words of type j. The two accounts are of course very similar to each other, to the extent that one might suspect them to be notational variants. The discussion on these issues has at least brought to light that certain things are more eas-

ily expressed in one framework than the other. For instance, it is easier to restrict cophonologies to local variation: we can determine certain small sets of constraints that are adjacent in the overall ranking and state that cophonologies can only use such adjacent sets. It is more difficult to justify in an indexed constraint approach why we cannot have in the same language some constraint being very high in the hierarchy for one type of words and very low in the hierarchy for another type of words. On the other hand, this lack of restriction has (of course) also be seen as an advantage as it makes the model more flexible. (See Inkelas and Zoll 2007 for a comparison between the two approaches.)
The cophonology idea also shows up in Stratal OT, in the sense that this model also allows for more than one ranking of constraints within the same language. In the spirit of Lexical Phonology, however, Stratal OT links each cophonology not to an arbitrary set of lexical items, but to a specific morphological 'stratum'; furthermore Bermúdez-Otero (2012) points out that the differences in ranking between different strata are restricted by considerations of learnability. If we have two cophonologies that are too far apart, the one which comes later in the chain of strata will cover most of the processes that happen in the earlier stratum, making it implausible for the learner to acquire that earlier ranking in the first place.

8. Psychological realism

One common criticism of scholars outside the OT framework has been the lack of psychological realism of the framework in most of its forms. In particular Classic OT presents a picture that has no hope of being implemented in a finite machine and/or a human brain, for instance because it requires the generation of an infinite number of candidate analyses for any input. The response of OT practitioners to this criticism has not been one of moving towards more psychological realism in the sense of setting up models that e.g. show the errors, although such work can also be found, e.g. in the work of Boersma, such as Boersma and Van Leussen 2017 (and references mentioned there). However, the aim of this work seems to be mostly to show that OT analyses can be implemented efficiently, *not* to make a direct claim about what goes on in the mind or brain (see also Chapter *). There has been some discussion about the question whether OT is even computationally tractable (Idsardi 2006, Kornai 2006).
The generative heritage of OT shows most clearly in the fact that a quite a lot of attention has gone to theories of learning. Indeed, this may be considered one of the successes of OT compared to most, if not all, other theories of phonology; different from generative syntax, very few previous theories of phonology seem to have been very concerned with the question whether the proposed formal systems of variation were also learnable. Within OT, on the other hand, several learning algorithms have been proposed. Already in the early nineties, Bruce Tesar demonstrated formally how Classic OT grammars can be learned (the work is summarised in Tesar 2013). Alternatives were also proposed relatively soon, based on extensions of the classic theory; Boersma and Hayes (2001) is the most prominent on of those. Next to explaining the actual typology of sound systems in the world, one of the main objectives of practitioners of the framework seems to be to explain how phonological systems can be learned (see also Chapter *).
Also in this case, the interest seems to be mostly in the so-called 'logical problem' of language acquisition, which aims to set up a formal system that can extract a

grammar from a representative sample of data, without too much attention to the actual order in which the grammar is learned or giving a realistic view on the kinds of errors that are made. As far as I am aware, the literature has very rarely made connections to the literature on the acquisition process among human learners. On the other hand, the literature on learning does connect to e.g. the logical-mathematical work of Prince (2002b, and subsequent work), which explores properties of the OT system by extracting the logic of ranking arguments and their entailments.

9. Applications to other fields

Since OT is a theory of variation and computation, it is not necessarily a theory of phonology only. As a matter of fact, there is no inherent reason to restrict its application to linguistics (although I do not know of serious application to other cognitive domains; see however Dresher 1996, Giannakopoulos 2001, Othero 2009). In principle, OT tableaux should be applicable to any kind of situation in which there are potentially conflicting demands on the output of a computation. As such, it should at least be applicable to other levels of grammatical analysis, such as phonetics, morphology, semantics and syntax; or otherwise, there should be a serious discussion of why those systems would work differently.
The theory has indeed been applied with considerable success to semantics and pragmatics, albeit usually in a slightly modified form (De Hoop and De Swart 2000, Legendre et al. 2016): so-called 'bidirectional OT', in which the evaluation does not compare an unchangeable input to some output, but in which two forms are evaluated at the same time (roughly corresponding to 'form' and 'meaning') and each can be changed to get the perfect match. This solves an interesting conceptual problem that Classic OT shares with other generative theories: that it seems to exclusively model production and not perception. Some work has shown that OT can itself be implemented in game theory and it is my impression that these game-theoretic analyses have become more popular over the years.
A similar approach is taken in the model of Boersma and Hamann (as in Hamann 2014 or Boersma and Van Leussen 2017) that combines phonology with phonetics; also this model is bidirectional in the sense that it can calculate concrete phonetic (acoustic and articulatory) representations from an abstract phonology and vice versa. The Boersma/Hamann model is, as far as I am aware, the only serious application of OT to phonetics.
It is quite striking that OT has never become very popular in syntax. There have been some early attempts at using the framework in syntax, collected for instance in Legendre at al. (2001), Broekhuis and Vogel (2009), as well as in the Legendre et al. (2016) volume mentioned above, but the framework did not achieve anything like the popularity in syntax as it did in phonology. The Rutgers Optimality Archive is mostly an archive of phonological papers. Within morphology, the framework seems to have worked mostly in the interface with phonology (such as Prosodic Morphology), but rarely outside it, although Trommer (2001) is an attempt to combine OT with Distributed Morphology (Halle and Marantz 1993). The majority of work in DM remains to assume phonological (and morphological) rules.
It is difficult to think of any explanation of this difference between phonology on the one hand and (morpho)syntax on the other which does not at least take into

account some component of sociology, taking into account the status of somebody like Chomsky in syntax (who as we saw opposed the theory in Chomsky 1995) and the status of Smolensky, Prince and McCarthy in phonology. It is difficult to think of anything inherent in phonological or syntactic computation and variation which would make one more suitable to a ranked constraint analysis than the other. Similarly, one could wonder why there is no real 'minimalist' phonology (with a few exceptions such as Samuels 2009). If there is such a difference between the two domains, that itself would be worthy of precise argumentation, but I know no real attempts to make a principled distinction between the domains from which it would follow that one works e.g. with binary parameters, and the other with ranked constraints.
Obviously, OT has never been the only framework within phonology either. My impression is that most opposition against OT in its heydays came from American scholars directly or indirectly related to MIT or Harvard, and in Europe from scholars connected to GP. Interestingly, both schools were interested in keeping a connection to syntactic theory, albeit in radically different, and maybe even opposing, ways. That might also point to a sociological factor playing a role, although that observation in itself of course has nothing to say about the correctness of either the OT enterprise or about its opponents.

10. Conclusion

Optimality Theory has without a doubt been the dominant framework of phonology for the past 25 years. It became such almost immediately when the first manuscript started circulating, and it has not been replaced by another theoretical framework in that role until today. This has as an advantage that many aspect of the theory were explored, so that we can quite safely state that 25 years after the first manuscript started circulating, we understand the theory quite well: we know what it can do and we know many of it limitations. It would probably surprise many scholars if somebody would come up with a consequence of OT that has not been foreseen by now. It has also importantly functioned as a lingua franca for analyses that in other circumstances would have been presented in a different way; a side effect of this is that many changes to the analysis were proposed to Classic OT, although very few of them became mainstream and hardly any served to restrict the power of the theory. This means that the term OT by now refers to a family of theories, with rather loose boundaries.
Although such things are hard to quantify, it seems to me that currently OT is slowly losing its grip on the field: fewer papers use the lingua franca in conferences and in journals. At the same time, scholars express the feeling that the field is falling apart:

> Phonology is changing rapidly... Some phonologists collect the evidence for their theories using introspection, fieldwork and descriptive grammars, while others trust only quantitatively robust experimentation or corpus data. Some test phonological theories computationally... whereas others prefer to compare theories on conceptual grounds.... As the field grows and diversifies, it is becoming harder for phonologists to talk to each other, for who can be a computer scientist, phonetician, neurolinguist and expert in adjacent fields such as

> morphology and syntax at the same time as having a command of the extensive literature on phonology-internal argumentation and phonological typology? (Gouskova 2013:173; see also Hyman 2015)

I believe that many of the tendencies that Gouskova describes were already at work in 1993. In some sense, the existence of the common notational system that was OT covered it and gave at least some sense of unity to the field. But if we look back at the period where OT was introduced, one can see that there was already quite a lot of difference of opinion on issues of the phonetic grounding, the importance of corpus research or laboratory experiments, the need to assign relatively abstract templates to syllable structure. One could argue that OT made claims about aspects of the theory that were relatively 'harmless' in this respect, while they were obviously important to almost everybody: the mapping of different representations onto each other, the way in which languages can vary. There were very few proposals on the market, and OT gave a formal tool that was not too difficult to learn and then could be applied to almost every problem, in spite of the many differences of opinion.
This may the reason why attempts to set up a theory of constraints have not been successful: such a theory can only be developed on the basis of more consensus of what the stuff of phonology really is: whether or not for instance the statistical distribution of variants in a corpus is one of the responsibilities of phonology, or whether historically grown patterns that are not productive are.
OT did help shed a lot of clarity on issues for which it was designed, but as it never pretended to be a theory of phonology as a whole, it would be unfair to require it to have resolved the many fundamental problems there are in the field.

References

Anttila, Arto. 2007. Variation and optionality. In Paul de Lacy (ed.), *The Cambridge Handbook of Phonology*. Cambridge, UK: Cambridge University Press. Pp. 519-536.

Archangeli, Diane, and Douglas Pulleyblank. 1994. *Grounded Phonology.* Cambridge, MA: The MIT Press.

Ashley, Karen C., Lynette Disch, David C. Ford, Erin MacSaveny, Steve Parker, Carla Unseth, Angela M. Williams, Rachel Wong, and Brendon Yoder. 2010. How many constraints are there? A preliminary inventory of OT phonological constraints. Manuscript, Graduate Institute of Applied Linguistics, Dallas. ROA 1079.

Bermúdez-Otero, Ricardo. 2012. The Architecture of Grammar and the Division of Labour in Exponence. In Jochen Trommer (ed), *The Morphology and Phonology of Exponence* (Oxford Studies in Theoretical Linguistics 41). Oxford: Oxford University Press. Pp. 8-83.

Bermúdez-Otero, Ricardo. 2018. Stratal Phonology. In S.J. Hannahs & Anna R. K. Bosch (eds), *The Routledge Handbook of Phonological Yheory.* Abingdon: Routledge. Pp. 100-134.

Blaho, Sylvia, Patrik Bye, and Martin Krämer (eds.) 2008. *Freedom of Analysis?* Berlin: Mouton de Gruyter.

Boersma, Paul and Bruce Hayes. 2001. Empirical Tests of the Gradual Learning Algorithm. *Linguistic Inquiry* 32: 45-86.

Boersma, Paul and Jan-Willem van Leussen. 2017. Efficient Evaluation and Learning in Multi-Level Parallel Constraint Grammars. *Linguistic Inquiry* 48: 349-388.
Booij, Geert. 2018. Phonology and the Dutch-Polish connection: a personal memoir. Manuscript, University of Leiden.
Brasoveanu, A. & Alan Prince. *Natural Language and Linguistic Theory* 29: 3. https://doi.org/10.1007/s11049-010-9103-3
Broekhuis, Hans, and Ralf Vogel (eds.) *Optimality Theory and Minimalism: Interface Theories*. Potsdam: Universitätsverlag Potsdam.
Burzio, Luigi. 1994. *Principles of English Stress.* Cambridge, UK: Cambridge Studies in Linguistics.
Cavirani, Edoardo. 2015. Brief history of the concept of constraint in Generative Phonology. Manuscript, Meertens Institute. http://www.academia.edu/1727433/Brief_history_of_the_concept_of_constraint_in_Generative_Phonology
Cedergren, H.; and D. Sankoff. 1974. Variable Rules: Performance as a Statistical Reflection of Competence. *Language*, 50, 333-355
Chomsky, Noam. 1965. *Aspects of the Theory of Syntax*. Cambridge, MA. MIT Press.
Chomsky, Noam. 1979. *Morphophonemics of Modern Hebrew.* New York & London: Garland.
Chomsky, Noam, and Morris Halle. 1968. *Sound Patterns of English.* Cambridge, Mass: MIT Press.
Coetzee, Andries. 2016. A comprehensive model of phonological variation: grammatical and non-grammatical factors in variable nasal place assimilation. *Phonology* 33:211-246.
Crowhurst, Megan J. 2011. Constraint Conjunction. In: M. van Oostendorp, C. Ewen, B. Hume and K. Rice (eds.) *The Blackwell Companion to Phonology.* Oxford: Blackwell. Pp. 1672-1694.
de Lacy, Paul (ed.). 2007). *The Cambridge Handbook of Phonology*. Cambridge University Press
Dresher, Elan. 1996. The Rise of Optimality Theory in First Century Palestine. *GLOT International* 2.1/2: 8.
Giannakopoulos, Andreas P. 2001. Frog Leaps and Human Noises. An Optimality Theory Approach to Cultural Change. Institute for Logic, Language and Computation, Universiteit van Amsterdam. Technical report.
Goldrick, M. 2001. Turbid Output Representations and the Unity of Opacity. in M. Hirotani, A. Coetzee, N. Hall, and J.-Y. Kim (eds.), *Proceedings of the Northeast Linguistics Society (NELS)* 30. Amherst, MA: GLSA. Volume I, pp. 231-245.
Goldwater, Sharon and Mark Johnson. 2003. Learning OT Constraint Rankings Using a Maximum Entropy Model. *Proceedings of the Workshop on Variation within Optimality Theory. Stockholm University, 2003*. Pp. 176–199.
Gouskova, Maria. 2013. Review of Marc van Oostendorp, Colin J. Ewen, Elizabeth Hume & Keren Rice (eds). 2011. *The Blackwell Companion to Phonology. Malden: Mass.:Wiley-Blackwell. Phonology* 30: 173- 179.
Green, Antony D. 2004. Opacity in Tiberian Hebrew: Morphology, Not Phonology. *ZAS Papers in Linguistics* 37: 37–70
Hamann, S. 2014. Phonological changes. In C. Bowern, & B. Evans (Eds.), *The Routledge handbook of historical linguistics*. (Routledge handbooks in linguistics). New York and London: Routledge, pp. 249-263.

Halle, Morris. 1957. In Defense of the Number Two. In Ernst Pulgram (ed.), *Studies Presented to Joshua Whatmough on his Sixtieth Birthday*. The Hague: Mouton. Pp. 65–72.
Halle, M. 1959. *The Sound Patterns of Russian*, The Hague: Mouton de Gruyter.
Halle, Morris & Alec Marantz. 1993. Distributed Morphology and the Pieces of Inflection. In Kenneth Hale and S. Jay Keyser (eds.) *The View from Building 20*, Cambridge: The MIT Press. Pp. 111–176.
Hayes, Bruce and Colin Wilson 2008. A Maximum Entropy Model of Phonotactics and Phonotactic Learning. *Linguistic Inquiry* 39: 379-440.
Hoop, H. de and H. de Swart. 2000. Optimality Theoretic Semantics. *Linguistics and Philosophy* 24: 1-32.
Hyman, Larry. 2015. Why Underlying Representations? Henry Sweet Lecture • Annual Meeting of the Linguistics Association of Great Britain (LAGB), September 16, 2015. University College, London.
Idsardi, William. J. 1997. Phonological Derivations and Historical Changes in Hebrew Spirantization. In Iggy Roca (ed.) *Derivations and Constraints in Phonology.* Oxford: Clarendon Press. Pp. 84-102.
Idsardi, William J. 2006. A Simple Proof that Optimality Theory is Computationally Intractable. *Linguistic Inquiry* 37: 271-275.
Inkelas, Sharon & Cheryl Zoll. 2005. *Reduplication*. Cambridge University Press.
Inkelas, Sharon & Cheryl Zoll. 2007. Is Grammar Dependence Real? A Comparison between Cophonological and Indexed Constraint Approaches to Morphologically Conditioned Phonology. *Linguistics* 45: 133-171.
Kager, René. 1995. English Stress. Re-Inventing the Paradigm. *GLOT International* 9/10: 19-21.
Kager, René. 1999. *Optimality Theory.* Cambridge, UK: Cambridge University Press.
Kiparsky, Paul. 1982. From Cyclic to Lexical Phonology. In Harry van der Hulst and Norval Smith (eds.) *The Structure of Phonological Representations, vol. I.* Dordrecht: Foris Publications. Pp. 131–75.
Kornai, Andras. 2006. Is OT NP-hard?. Manuscript. ROA-838.
Krämer, Martin. 2008. *Vowel Harmony and Correspondence.* Berlin: Mouton de Gruyter.
Legendre, Géraldine, Jane Grimshaw and Sten Vikner. 2001. *Optimality-Theoretic Syntax*. Cambridge, MA: The MIT Press.
Legendre, Géraldine, Yoshiro Miyata and Paul Smolensky. 1990. Harmonic Grammar: A Formal Multi-Level Connectionist Theory of Linguistic Well-Formedness: Theoretical Foundations. *Proceedings of the twelfth annual conference of the Cognitive Science Societ*. Cambridge, MA: Lawrence Erlbaum. Pp. 388–395.
Legendre, Géraldine, Michael T. Putnam, Henriëtte de Swart and Erin Zaroukian. 2016. *Optimality Theoretic Syntax, Semantics, and Pragmatics From Uni- to Bidirectional Optimization*. Oxford: Oxford University Press.
Magri, Giorgi, and Arto Antilla. 2018. Does MaxEnt overgenerate? Implicational universals in Maximum Entropy grammar. In Gallagher, Gillian and Gouskova, Maria and Sora, Yin (eds.) *AMP 2017: Proceedings of the 2017 Annual Meeting on Phonology.*.
McCarthy, John J. 1988. Feature Geometry and Dependency. A Review. *Phonetica* 38: 23-37.

McCarthy, John J. 1999. Sympathy and Phonological Opacity. *Phonology* 16: 331-399.
McCarthy, John J. 2003. Comparative Markedness. *Theoretical Linguistics* 29: 1-51.
McCarthy, John J. 2007. *Hidden Generalizations: Phonological Opacity in Optimality Theory*. London: Equinox Publishing.
McCarthy, John J. 2010a. An Introduction to Harmonic Serialism. *Language and Linguistics Compass*, 4(10):1001–1018.
McCarthy, John J. 2010b. Studying Gen. *Journal of the Phonetic Society of Japan*.
McCarthy, John J. and Alves, Ubiratã K. 2017. Optimality Theory: an Interview with John McCarthy. *ReVEL*, vol. 15, n. 28.
McCarthy, John J. and Alan Prince. 1993. Prosodic Morphology I: Constraint Interaction and Satisfaction. Linguistics Department Faculty Publication Series. 14.
McCarthy, John and Prince, Alan. 1995. Faithfulness & reduplicative identity. In Beckman, Jill, Laura Walsh Dickey, Laura, and Suzanne Urbanczyk (eds.) *University of Massachusetts Occasional Papers in Linguistics 18: Papers in Optimality Theory.* Amherst: Graduate Linguistics Student Assocation. Pp. 249-384.
McCarthy, John J. and Alan Prince. 1996. Prosodic Morphology. In: John Goldsmith (ed.) *The Handbook of Phonological Theory.* Oxford: Blackwell.
Nagy, Naomi. 2013. Phonology and Sociolinguistics. In Robert Bayley, Richard Cameron, and Ceil Lucas (eds.) *The Oxford Handbook of Sociolinguistics*. Oxford: Oxford University Press.
Nespor, Marina, and Irene Vogel. 1986. *Prosodic Phonology.* Dordrecht: Foris.
Odden, David. 2011. Rules vs. Constraints. In John Goldsmith, Jason Riggle and Alan Yu (eds.) *The Handbook of Phonological Theory.* Oxford: Wiley.
Orgun, Orhan. 1996. *Sign-based Morphology and Phonology: with Special Attention to Optimality Theory*. PhD dissertation, UC-Berkeley.
Othero, Gabriel de Avila. 2009. Optimality Theory and the Three Laws of Robotics. Manuscript, Universidade Federal do Rio Grande do Sul.
Pater, Joe. 2016. Universal Grammar with Weighted Constraints. In John McCarthy and Joe Pater (eds.) *Harmonic Grammar and Harmonic Serialism*. London: Equinox Press. Pp. 1-46.
Prince, Alan. 2002a. Arguing Optimality. In Andries Coetzee, Angela Carpenter and Paul de Lacy (eds). *Papers in Optimality Theory II*. GLSA, UMass.Amherst.
Prince, Alan. 2002b. Entailed Ranking Arguments. Manuscript, Rutgers University. ROA-500.
Prince, Alan, and Paul Smolensky. 1993. *Optimality Theory. Constraint Interaction in Generative Grammar*. Manuscript, Rutgers University and Johns Hopkins. Published in 2004 with Blackwell.
Revithiadou, A. 2007. Colored Turbid Accents and Containment: A Case Study From Lexical Stress. In Sylvia Blaho, Patrik Bye and Martin Krämer (eds.) *Freedom of Analysis?* Mouton de Gruyter, Berlin & New York. Pp. 149-174.
Rubach, Jerzy. 2003. Polish Palatalization in Derivational Optimality Theory. *Lingua* 113: 197–237.
Samuels, Bridget. 2009. *The Structure of Phonological Theory.* PhD Dissertation, Harvard.

Smolensky, Paul. 2006. Harmony in Linguistic Cognition. *Cognitive Science* 30: 779–801

Tesar, Bruce. 2013. *Output-Driven Phonology.* Cambridge: Cambridge University Press.

Torres-Tamarit, Francesc. 2015. Length and Voicing in Friulian and Milanese: Or Why Rule-Free Derivations Are Needed. *Natural Language and Linguistic Theory* 33. 1351–1386.

Trommer, Jochen. 2001. *Distributed Optimality*. PhD dissertation, Universität Potsdam.

Trommer, Jochen, and Eva Zimmermann. 2014. Generalised Mora Affixation and Quantity-Manipulating Morphology. *Phonology* 31: 463-510.

Walker, Rachel. 2000. Long-Distance Consonantal Identity Effects. In Roger Billerey and Brook Danielle Lillehaugen (eds.) *Proceedings of the 19th West Coast Conference on Formal Linguistics*. Somerville, MA: Cascadilla Press. Pp. 532-545.

Zimmermann, Eva. 2017. *Morphological Length and Prosodically Defective Morphemes.* Oxford: Oxford University Press.