CHAPTER 21

OPTIMALITY THEORY IN PHONOLOGY

MARIA GOUSKOVA

21.1 Introduction

The goal of this chapter is to overview Optimality Theory (OT, Prince and Smolensky 2004) as applied to phonology.[1] OT is a theory of constraint interaction in grammar, which aims to solve a couple of related problems that have confronted generative phonological theory since its earliest days. The first problem is conspiracies: in some languages, there is a constraint that seems to be satisfied in a variety of ways, as if the rules conspire to achieve a single target. The second problem is soft universals: unrelated languages show evidence of the same or similar constraints, but the constraints do not seem to hold in all languages. Moreover, as in single-language conspiracies, the way these constraints are satisfied may differ from

I'd like to think Amanda Dye, Joe Pater, Jen Smith, Jim Wood, and the editors for feedback.

[1] Because of space limitations, the review is necessarily incomplete, but there is no shortage of other article-length treatments of OT (Prince and Smolensky 2002, McCarthy 2007*b*, Tesar et al. 1999, Smolensky et al. 2006) or book-length overviews (McCarthy 2008, 2002, Kager 1999, Archangeli and Langendoen 1997, Prince and Smolensky 2004). In addition to the book-length collections of phonology papers in OT cited in the body of the chapter, there are other general (McCarthy 2003*a*) and topical collections (Lombardi 2001*a*, Roca 1997). Last but not least, there is an extensive free online archive of papers in and about OT, the Rutgers Optimality Archive, at http://roa.rutgers.edu.

language to language. OT addresses both problems by introducing the assumption that constraints are universal but rankable and violable. This simple assumption has many surprising consequences, which have been fruitfully pursued in the fifteen years since the advent of the theory.

Conspiracies were discovered by Kisseberth (1970*a*), who describes several rules in Yawelmani that are united functionally but couldn't be unified formally in the theory of the time (*The Sound Pattern of English*/SPE, Chomsky and Halle 1968). Another example comes from Tonkawa (Kisseberth 1970*b*, McCarthy 1986). Tonkawa has a rule of vowel deletion, which happens to be blocked just in case it would create a geminate, or long consonant (see (1)).

(1) A constraint on vowel deletion in Tonkawa (Hoijer 1933)
 a. Vowel deletion between non-identical consonants
 /noto̲xo-n-oʔ/ notxonoʔ 'he is hoeing it' cf. notox 'hoe'
 /pice̲na-n-oʔ/ picnanoʔ 'he is cutting it' cf. picen 'castrated steer'
 b. No vowel deletion if surrounding consonants are identical
 /hewa̲wa-n-oʔ/ hewa̲wanoʔ 'he is dead' *hewwanoʔ
 /ham'a̲m'a-n-oʔ/ ham'a̲m'anoʔ 'he is burning' *ham'm'anoʔ

Another rule of vowel deletion, which deletes the stem-final vowel in compound formation, may apply even between identical consonants, but what surfaces is a single short consonant (see (2)). When two identical consonants are brought together by morpheme concatenation, one of them also deletes (see (3)). In this conspiracy, the rules of vowel deletion and consonant shortening work together to avoid geminates:

(2) Tonkawa compound vowel deletion and geminate simplification (McCarthy 1986: 225)
 /taʔane̲-nosʔo:yta-/ taʔanosʔo:yta- 'to stretch (e.g., a rope)'
 /yakona̲-nacaka-/ yakonacaka- 'to kill (him) with a blow of fist'
 /yakexe̲-xakana-/ yakexakana- 'to push (it) down hard'

(3) Tonkawa morpheme concatenation and geminate simplification (Hoijer 1949)
 /nes-so:la-/ neso:la- 'to spill (it)' cf. nes-kapa- 'to shut a door'
 /ʔey-yace-/ ʔeyace- 'to catch, capture (them)' cf. ʔey-pake- 'to slice it'

This anti-geminate prohibition applies not only to derived geminates: there are no geminates even morpheme-internally in the language, so morphemes like **hewwa*- are absent.

(4) A constraint on Tonkawa morphemes
 $* \ldots C_i C_j \ldots$, where i = j

An insightful analysis of Tonkawa would explain why the geminate prohibition holds both in derived and in underived sequences. It should also capture the obvious connection between this prohibition blocking vowel deletion and triggering consonant shortening. Yet pre-OT treatments had to explain such generalizations through two separate mechanisms. Restrictions on underived sequences were handled through Morpheme Structure Constraints, which held at the underlying level. Restrictions on derivations were stated as part of the rule's context, or else put into a separate "derivational constraint" whose interaction with the rule it blocked was never fully formalized (Kisseberth 1970*b*). It was likewise impossible to draw any connection between such constraints and the rules they trigger, as in the case of consonant shortening. Any similarity between a Morpheme Structure Constraint and a condition on some rule's application was purely coincidental, creating a redundancy known as the *duplication problem* (see McCarthy 2002, §2.1 for discussion). OT avoids the duplication problem by assuming that the constraint against geminates applies only to surface forms, prohibiting both derived and underlying geminates. It blocks vowel deletion because it overrides the constraints that make vowel deletion necessary in the first place. OT can also make sense of the way the geminate prohibition apparently triggers consonant shortening because constraints in OT can interact to compel such alternations.

The prohibition against geminates is a kind of soft universal. It blocks vowel deletion in many unrelated languages such as Afar and Tiberian Hebrew (McCarthy 1986). As Odden (1988) shows, however, there are languages where this is not the case.[2] Even in languages that freely violate the constraint against geminates, there is often evidence that they are disfavored. In a theory without violable constraints, counterexamples to a purportedly universal constraint immediately put its validity into question. Existing solutions are all somewhat unsatisfying—for example, sometimes it is posited that the principles hold at different levels of derivation in different languages, or they are treated as parameters with language-specific settings. Nevertheless, theories with inviolable constraints have no way of capturing the intuition that the same constraint seems to be at work even though it appears to be violable. On the other hand, soft universals are unsurprising for OT, since OT constraints are violable and universal: languages may either satisfy constraints or skirt them altogether. OT furthermore predicts that constraints can be satisfied partially even if they are generally violated in a language. This kind of interaction, of which there is ample evidence in work on OT, is known as the Emergence of the Unmarked (McCarthy and Prince 1994).

By assuming that constraints are universal and violable, OT suggests a natural theory of typology. In the strongest version of the theory, constraints are universal, and any reordering of them should produce a plausible grammar. This simple premise makes for an easily falsifiable theory of phonological grammar, and it has many interesting consequences. For example, OT allows for a principled approach

[2] There are still others that have an almost reverse pattern of deleting a vowel only between identical consonants; for a recent OT account and typology, see Bakovic (2005).

to the problem of variation. Variation can be framed as two or more different grammars that coexist within a speaker or a community. Since OT explicitly formalizes the notion of differences between grammars, it can account for variation with a few modest extensions. The problem of learning can also be understood in similar terms: how does a learner arrive at the right grammar when starting out with an incorrect one?

The chapter is organized as follows. Section 21.2 describes the architecture of the theory, including its basic components (§21.2.1–21.2.3.2) and approach to typology (§21.2.3.3). Section 21.3 addresses the status of the lexicon in OT. Sections 21.4 and 21.5 describe some work on learnability, acquisition, and variation. Section 21.6 concludes.

21.2 Architecture

An OT grammar has three components. Con is the component that defines the set of universal violable constraints. Gen is the component where output candidate parses are generated based on input forms. Eval is the component that selects an optimal output from the set of alternative candidates, given a language-specific hierarchical ordering of Con, $\mathcal{H}$. The path from the input to the output is charted in (5). Even though versions of OT may differ from each other in the details of how Gen and Eval are implemented, most work in OT assumes something like (5).

(5) OT: the organization of the grammar

/input/	→	Gen	→	Eval($\mathcal{H}_{CON}$)	→	[output]
		candidate 1				
		candidate 2				
		candidate 3				
		...				

Each of these components is examined in turn in the following sections, starting with Eval (§21.2.1) and moving on to Gen (§21.2.2) and Con (§21.2.3.2). The focus throughout will be on phonological issues and applications, set in the most widely accepted version of the theory known as "classic" or "parallel" OT (Prince and Smolensky 2004, McCarthy and Prince 1995).[3] The main principles of this version of the theory are outlined in Prince and Smolensky (2004); since it is impossible to do justice to all the work done on the theory since 1993, interested readers are referred to the various works cited along the way for more recent developments.

[3] Several proposals depart from this architecture by assuming a constrained or modified Gen (McCarthy 2007*a*) or multiple serially ordered evaluations by the constraint hierarchy (Stratal OT; Kiparsky to appear). Some versions include an additional component that further filters the output of Eval (Orgun and Sprouse 1999, de Lacy 2007).

21.2.1 Eval

One of the defining features of OT is competition between candidates, and this competition is resolved in the Eval module. Eval considers candidates in pairwise comparisons based on their relative performance with respect to the constraint hierarchy. For an informal example, consider the Tonkawa pattern described above. Given the input /hewawa-n-oʔ/, the grammar has a choice between a candidate that deletes the vowel, *[hewwanoʔ], and one that does not—the actually attested [hewawanoʔ]. Vowel deletion is required by the general phonology of Tonkawa, so not deleting violates a constraint, but deleting creates a geminate, which also violates a constraint. The fact that deletion is blocked suggests that "NoGeminates" dominates "DeleteVowel". (These constraints are just placeholders for now; I return to the actual constraints at work in §21.2.3.2.)

More abstractly, two constraints disagree on a pair of candidates when one of the constraints favors the candidate that the other constraint disprefers. In the schematic example (6), Constraint 1 and Constraint 2 disagree in just that way on outputs 1 and 2 (an asterisk in a constraint's column indicates that the candidate violates it). Output 1 is more *harmonic* than output 2 with respect to Constraint 1, and the opposite is true for Constraint 2. We infer that Constraint 1 dominates Constraint 2 because output 1 is the surface form. Its *optimal* status is marked here with ☞:

(6) Candidate evaluation by constraints

	/input/	Constraint 1	Constraint 2
a.☞	output 1		*
b.	output 2	*!	

Proof of constraint ranking is given in a *tableau* such as (6), or in the more compact format of a comparative tableau (see (7)). In a comparative tableau, the optimum (given first) is paired with a loser. Since, as shown above, Constraint 1 prefers the winner of the competition, a W appears in its column. An L in the column of Constraint 2 indicates that it prefers the loser.

(7) A comparative tableau

/input/	Constraint 1	Constraint 2
output 1_W~output 2_L	W	L

The optimum is the candidate that best satisfies the highest-ranked constraint that distinguishes it from other candidates. It is often the case that for a given winner~loser pair, some constraint does not distinguish the candidates, either because both violate the constraint or because both satisfy it to the same extent. In such cases, the constraint does not participate in determining the outcome, and the optimum must be chosen by some other constraint. Thus, if each of the outputs in (7) violated some Constraint 3 three times, it would not affect the optimality of output 1 even if that constraint happened to be ranked above Constraint 1. What matters is not absolute satisfaction of constraints but rather comparative performance.

This model of Eval is assumed in most OT work, both in phonology and in other areas (though there are other theories of Eval, e.g., Wilson 2001). Notably, this usually ensures a unique winner for any competition, since there is usually at least some constraint that distinguishes even very similar candidates. This architecture of Eval must be changed, however, in order for the theory to produce variation or optionality; some proposed modifications are discussed in §21.5.

21.2.2 Gen

Gen is the component of an OT grammar that generates the competing candidates from which the output is chosen. Since there is no primitive notion of rules or transformations in OT, it falls to Gen to produce a wide enough range of forms that would cover the range of phonological operations, though it even goes beyond that, as we will see shortly. Basically, the job of Gen is to improvise on the input. In phonology, Gen can map an input to an output more or less without changes, or it can modify the input. There are several ways to render an input faithfully. An input like /patra/ can be syllabified as either [pa.tra] or [pat.ra]. Gen can also manipulate the input by changing distinctive features, deleting or inserting segments, assigning stress, and so on. In addition to assigning phonological structure, Gen tracks how each input is mapped to each output candidate by positing a relation between elements of the input and the elements of the output. Thus, each candidate is not just an output form but also a mapping from the input. This input–output relation is typically formalized in Correspondence Theory (McCarthy and Prince 1995). Some examples of candidates for a hypothetical input /bak/ are given below.

(8) Some candidates emitted by the phonological Gen

input	*output*	*comments*
/bak/	bák	fully faithful candidate
	bá.ki̱	epenthesis, stress on the first syllable
	ba.kí̱	epenthesis, stress on the second syllable
	bá	deletion
	i̱.bá.ki̱	double epenthesis
	káb	metathesis (reordering)
	vák	feature change (frication)
	pák	feature change (devoicing)
	. . .	

An important property of Gen in traditional OT is that it is unconstrained by phonological well-formedness principles and knows nothing about constraint satisfaction. Instead, Gen is usually assumed to have *Freedom of Analysis* (McCarthy and Prince 1993*b*): it can modify the input in all sorts of ways, many of which may seem absurd to a linguist. For example, given the hypothetical input /pata/, Gen is allowed to map it to the sensible [(pá.ta)] or [(pa.tá)],[4] but also to [(pá)(tà)], [(pá.ta)ʔə], [(pán.da)], [(á.tap)], [(p(á)t.a.)], and so on. This somewhat counterintuitive feature of OT stems from the imperative of decoupling grammatical operations from the constraints that determine what surfaces. The operations themselves are of relatively little interest, but well-formedness is paramount, and it is a matter for constraints to sort out.

This rich array of candidates emitted by Gen has been argued to be problematic. At least in part, it contributes to the so-called too-many-solutions problem (a term due to Steriade 2001): there are many conceivable ways of avoiding certain marked structures, yet many seem to be unattested cross-linguistically. One simple example is avoidance of final voiced obstruents. While word-final devoicing (/pad/ → [pat]) is extremely common, vowel epenthesis (/pad/ → [pa.di]) and consonant deletion (/pad/ → [pa]) appear to be unattested. This is all the more puzzling since laryngeal features are different in this respect from place features. Lombardi (2001*b*) proposes a solution that relies in part on a revision of Gen and in part on certain assumptions about Con, but similar problems arise in other areas, so there is no universal solution. It should be noted that the too-many-solutions problem is an issue not just for OT but for any theory of phonology that aims to account for typology: if the right theory of phonology is rule-based, there is still a question of why certain rules seem to be ubiquitous and others don't seem to occur, which was never satisfactorily solved.

Some recent work, however, challenges Freedom of Analysis (see the various contributions to Blaho et al. 2007). In McCarthy's (2007*a*) OT with Candidate Chains, candidates are generated in incremental steps, and each step is checked against the constraint hierarchy to ensure that the change increases well-formedness. McCarthy argues that this revision can address the too-many-solutions problem; see also Wilson (2001). Interesting results may also be obtained by changing the way Gen manipulates segmental, syllabic, and metrical structure, and there are arguments that this is actually a necessary restriction on the theory (Morén 2007, Smith 2007, Rice 2007, Lombardi 2001*b*).

A typical example of what could be at stake comes from the realm of metrical foot structure, discussed by Rice (2007). Most modern work on metrical stress theory assumes that feet are binary pairings of weak and strong elements, drawn from syllables or moras (Prince 1985, Bakovic 1998). Other theories admit feet

[4] Square brackets signify prosodic words, round brackets delimit feet, and dots show boundaries between syllables. Acute accents stand for primary word stress, and grave ones for secondary stress.

978–0–19–954400–4 21-Heine-C21-drv Heine-Narrog (Typeset by Spi, Chennai) 538 of 1025 June 30, 2009 16:19

with more than two syllables (Kenstowicz 1996, Hayes 1995). There is one area where such feet could offer an analytical advantage: stress in languages such as Cayuvava, which falls on every third syllable (see Halle and Vergnaud 1987, Levin 1988, and cf. Elenbaas and Kager 1999). As Rice points out, any theory of ternary stress couched in OT must confront two separate questions. The first is whether ternary feet are necessary for analyzing ternary stress. The answer is no—analyses in terms of binary feet are possible. The second question is whether feet of three syllables and larger should be ruled out universally, and if so, how. Some properties of feet can be derived through constraint interaction alone, but others may need to be stipulated to be properties of Gen. If ternary feet are ruled out in Gen, there needs to be a principled theory behind such a prohibition, which is at present lacking.

OT is in principle compatible with different assumptions about phonological structure, since it is a theory of the architecture of the grammar rather than of phonological representations. It is, however, impossible to discuss properties of Gen without making specific theoretical assumptions about substantive properties of phonological theory, i.e., feet, syllables, and features. This is an area of ongoing and future work.

One final issue that is relevant here is the lack of derivations in parallel OT. Unlike SPE and much work in syntactic theory in the generative tradition, OT has just a single-step mapping from the input to the output instead of incremental derivational steps. This addresses certain problems such as top-down interactions between different levels of structure (see Prince and Smolensky (2004) on Tongan, for example), but it introduces another problem, namely a difficulty with certain types of opaque interactions. The problem is too complex to review here in any detail, but there are several proposed solutions. They include reintroducing derivational levels (Kiparsky to appear), special candidates that mimic derivational stages (McCarthy 2003*c*), and, finally and most relevantly, a ground-up revision of Gen that actually includes whole derivations as candidates (McCarthy 2007*a*). The latter book includes a comprehensive overview of the issue and work both in OT and in other theories.

21.2.3 Con and Factorial Typology

In this section, I overview some aspects of the constraint component Con that are assumed in much modern work in phonology. The prevailing theme here is that Con is not homogeneous, and it is not an arbitrary list of ad hoc constraints; rather, it has elaborate structure. At the very least, constraints are classified into markedness, faithfulness, and interface constraints. Within each of these types, there are subfamililes, grouped based on the way they relate to linguistic primitives. I discuss each type in turn. The last subsection deals with OT's approach to typology.

978–0–19–954400–4 21-Heine-C21-drv Heine-Narrog (Typeset by Spi, Chennai) 539 of 1025 June 30, 2009 16:19

21.2.3.1 *Internal structure of CON: constraint types*

Intuitively, markedness constraints ban elements that are structurally complex or in some sense difficult or disfavored. For example, in the realm of syllable structure, complex onsets such as [pra] are marked compared to simplex onsets such as [pa]; hence, there is a markedness constraint *Complex against tautosyllabic consonant clusters. The definitional property of markedness constraints, however, is not that they ban difficult things but rather that they refer only to output structures. Markedness constraints are often but not always based on phonetic principles; they may also have formal origins. Constraints of both types are discussed in the following sections.

By contrast, faithfulness constraints govern disparities between two levels of representation. The most familiar levels are the input (underlying representation) and the output (surface representation); for example, the constraint Max (McCarthy and Prince 1995) requires every segment in the input to have a correspondent in the output. This requirement is violated by deletion (e.g., /pra/ → [pa]). Almost every kind of disparity[5] between input and output violates some faithfulness constraint: deletion, insertion, reordering/metathesis, featural changes, and other operations of Gen discussed in §21.2.2 all have associated faithfulness costs. Faithfulness constraints in phonology can mediate between segmental strings at other levels of representation, as well. Since the same kinds of disparity sometimes result from inexact copying in the domain of morphological reduplication, McCarthy and Prince (1995) propose a unified theory of faithfulness, according to which the same types of constraints mediate between input vs. output and reduplicative base vs. reduplicant copy. Benua (1997) extends the theory to apply between words that are related by morphological derivation, and other proposals have since extended faithfulness to other domains.

To illustrate the markedness–faithfulness distinction, consider Tonkawa vowel deletion. Recall that in Tonkawa, vowels delete between two non-identical consonants but not between identical ones. The first step in analyzing Tonkawa is explaining why vowel deletion happens at all. Deletion creates a mismatch between the input and the output, which violates faithfulness. This means that some markedness constraint dominates faithfulness. The logic here is intuitively simple: given the way Eval works, there must be a tradeoff for violating faithfulness, and the only reason for a candidate to map unfaithfully is to become less marked.[6] Suppose the tradeoff is satisfying the requirement that stressed syllables be long, or heavy: if the vowel is deleted, the remaining consonant can close the preceding syllable.[7]

[5] Not all operations of Gen do, though. It is usually assumed that syllabification is not associated with faithfulness violations, since it is not contrastive. If a faithfulness constraint were violated by positing or changing syllable structure, then we would expect to see contrasts in syllable structure alone. See, for example, McCarthy 2003*c*.

[6] This property of OT is known as Harmonic Ascent (McCarthy 2002, Moreton 2003).

[7] A full analysis of Tonkawa vowel deletion along these lines is developed in Gouskova (2003).

This is shown in (9): the winner is unfaithful, since it deletes the second underlying vowel, and the loser is marked, since it has a stressed open syllable [nó...]. The markedness constraint Stressed=Heavy, which requires stressed syllables to be heavy, dominates the faithfulness constraint Max-V, which prohibits the deletion of vowels:

(9) Tonkawa: vowel deletion required between non-identical consonants

/notoxo-n-oʔ/	Stressed=Heavy	Max-V
nót.xo.nóʔ∼nó.t<u>o</u>.xo.nóʔ	W	L

Markedness and faithfulness constraints are often in conflict, but markedness constraints can also conflict with other markedness constraints. In Tonkawa, the constraint against geminates, NoGem, blocks the normal application of vowel deletion. This conflict is shown in (10). Here, the markedness constraints conflict with each other since the choice is between having a geminate on the one hand and having a stressed open syllable on the other. No-Gem actually agrees with Max-V on the candidates in 10. Thus, whether markedness and faithfulness constraints conflict really depends on the candidate and the constraints in question.[8]

(10) Tonkawa: vowel deletion blocked between identical consonants

/hewawa-n-oʔ/	NoGem	Stressed=Heavy	Max-V
hé.w<u>a</u>.wá.noʔ∼héw.wa.nóʔ	W	L	W

The third type of constraints in OT are interface constraints generated by Generalized Alignment (discussed in §21.2.3.2.1) and morpheme realization constraints, which require that morphological entities be realized as phonological content (see McCarthy and Wolf 2005 for recent discussion). These share some features with faithfulness, since they also mediate between two levels of structure, but instead of looking at the same type of structure instantiated in the input and the output, they require access to structures from different components of the grammar. A typical interface constraint will require a given edge of a phonological structure to coincide with some edge of some morphological structure. An example would be the requirement for a phonological word to end with a segment that belongs to a lexical/morphological word (Selkirk 1995, McCarthy and Prince 1993*a*). This requirement is violated by enclitics, which are prosodified with the preceding lexical word (e.g., English possessive *Mary'<u>s</u>*), and by inserted segments, which are assumed not to have any morphological affiliation at all (e.g., Lardil augmentation /kaŋ/ → [kaŋ<u>ka</u>], Prince and Smolensky 2004). Unlike the markedness/faithfulness dichotomy, such interface constraints are not an essential feature of the

[8] This analysis of Tonkawa vowel deletion has not yet addressed the problem brought up at the outset, namely the general absence of tautomorphemic geminates in Tonkawa, but I return to it in §21.5.

architecture of OT; indeed, they are very much a feature of a particular theory of the interface between morphosyntax and phonology (Selkirk and Shen 1990 and others).

Thus far, we've seen three types of constraints: markedness, faithfulness, and interface constraints. They are classified based on the structural levels they access in the process of evaluation. Constraints may further be classified based on the way they relate to linguistic primitives such as phonetic scales and atomic elements of phonological representation; this is the subject of the next section.

21.2.3.2 *Internal structure of CON: constraint schemata*

The content of the constraint component of the phonological grammar should be the most controversial aspect of the theory, since its typological predictions depend on how the constraints interact with each other under re-ranking. Unsurprisingly, a considerable effort in early OT work was devoted to discovering the constraints and working out their relationships to each other and to the substantive principles thought to underlie phonological patterns. One of the most productive lines of attack on this has been the development of constraint schemata, which define families of constraints based on how they are built from phonological primitives. The two best-known constraint schemata are Generalized Alignment (McCarthy and Prince 1993*a*) and Harmonic Alignment (Prince and Smolensky 2004). I discuss each in turn.

21.2.3.2.1 *Generalized Alignment* Generalized Alignment is a constraint schema proposed by McCarthy and Prince (1993*a*) in the context of a theory of edge effects. Constituent edges are often the domain of special phonology. Thus, stress is attracted to word edges: descriptions of stress patterns often make references to initial, final, and penultimate syllables. Which edge wins as a default is up to the language, but the orientation of stress toward word edges is a near-universal feature of stress (Hayes 1995). Similarly, languages may differ in whether they require prosodic word edges to coincide with lexical word edges or whether there can be mismatches between the two types of structures. Under Generalized Alignment, these sorts of observations are captured by constraints that require edges of particular domains to coincide with edges of other domains: metrical feet must be aligned to prosodic word edges, lexical words must be aligned to prosodic word edges, and so on. The basic ingredients of an alignment constraint are constituents to be aligned and the edge(s) that must coincide. McCarthy and Prince (1993*a*) define Generalized Alignment as follows:

(11) Generalized Alignment
Align (Cat1, Edge1, Cat2, Edge2)$=_{def}$

∀ Cat1 ∃ Cat2 such that Edge1 of Cat1 and Edge2 of Cat2 coincide.
Where Cat1, Cat2 ∈ PCat ∪ GCat
Edge1, Edge2 ∈{Right, Left}

In the original proposal, Generalized Alignment applied to prosodic (Selkirk 1978 and others) and morphological constituents, as shown in (12). It has since been extended to many other phonological structures and representational primitives, including subsegmental features (Kirchner 1993), tones (Myers 1994), metrical grids (Gordon 1999). Alignment is such a general formalism for constraint definitions that some have proposed to rethink even familiar constraints such as Onset and NoCoda in alignment terms: Onset requires simply that the left edge of a syllable must coincide with the left edge of a consonantal segment (McCarthy and Prince 1993*a*, Ito and Mester 1994). Similarly, Ito and Mester (1994) propose that the old CodaCond of Ito (1986) should be undersood as a family of alignment constraints that require certain features that are marked in coda position to be aligned to the beginning of a syllable rather than the end.

(12) Categories referenced by alignment

PCat = levels in the Prosodic Hierarchy	GCat = morphological constituents
ProsodicWord	MorphWord → Stem*
Foot	Stem → Stem, Affix
syllable	Stem → Root

An issue separate from which constituents to align is how edge alignment constraints assign violation marks. This is actually a problem that extends beyond alignment to all constraint evaluation. In the original proposals (Prince and Smolensky 2004, McCarthy and Prince 1993*a*), some alignment constraints were assumed to be violated gradiently: a single instance of a misaligned structure could incur more than one violation of the constraint, depending on degree of deviation from perfect alignment. McCarthy and Prince (1993*a*) assume that metrical foot alignment constraints such as All-Feet-Right "The right edge of each foot corresponds with the right edge of some prosodic word" assign a violation mark for each syllable that stands between the foot and the prosodic word edge. If there is more than one foot in a word, each foot's misalignment contributes to total violations:

(13) Gradient evaluation of edge alignment

	All-Feet-Right	*comments*
a. σσ(σ́σ)	✓	perfect alignment on the right
b. σ(σ́σ)σ	*	one foot, misaligned by one syllable
c. (σ́σ)σσ	**	one foot, misaligned by two syllables
d. (σ́σ)(σ́σ)σ	****	two feet, misaligned by one and three syllables

Gradient evaluation captures certain aspects of edge-sensitive phenomena very well: for example, it offers a straightforward analysis of antepenultimate stress (as in Macedonian, for example) as the resolution of a conflict between NonFinality and All-Feet-Right. Intuitively, in Macedonian, the foot is placed as close as possible to the right edge of the word, but not so close as to encompass the last syllable; since perfect right-alignment is impossible with NonFinality(Foot) ranked above All-Feet-Right, the next best option is chosen instead because alignment is violated only minimally. Unfortunately, this approach also appears to overgenerate by predicting certain unattested patterns in stress, infixation, and harmony systems (Kager 2001, McCarthy 2003*b*). This overgeneration is one of several arguments (McCarthy 2003*b*) for the claim that all constraints are categorical: a candidate should only incur multiple violations of a constraint if it has more than one instance of a structure that violates the constraint. As noted earlier, however, this controversy is separate from Generalized Alignment as a substantive theory of edge phonology.

21.2.3.2.2 *Harmonic Alignment* Harmonic Alignment is a theory of how phonetic and other extralinguistic scales are expressed in the grammar. In the most general form, Harmonic Alignment postulates that there is a relation between prominence and position: prominent positions are ideally filled with prominent elements, and non-prominent positions are filled with non-prominent ones. By now, Harmonic Alignment has been productively extended to generate phonological constraints on sonority-sensitive stress (Kenstowicz 1996), positional vowel reduction (Crosswhite 1999), and tone-stress interactions (de Lacy 2002*a*). Originally, however, Prince and Smolensky (2004) proposed Harmonic Alignment specifically to capture the well-known role of sonority in syllabification (see, for example, Clements 1990), and I'll discuss this application of it here. Prince and Smolensky observe that the prominent position of syllable nucleus is ideally filled with the most sonorous segment, i.e., a vowel, and the non-prominent position of syllable margin (i.e., onset) is ideally filled with an obstruent. Under their proposal for Harmonic Alignment, the position and prominence scales in (14) and (15) would combine to give two scales, one of which defines harmony for nuclei, and the other for margins. Harmonic Alignment is formulated as follows:

(14) Syllable position
Nucleus > Margin (Onset)

(15) Sonority scale
Vowels > Liquids > Nasals > Obstruents

(16) Harmonic Alignment (Prince and Smolensky 2004)
Given binary dimension D_1 with a scale X > Y on its elements {X, Y}, and another dimension D_2 with a scale a > b > . . . > z on its elements, the *harmonic alignment* of D_1 and D_2 is the pair of harmony scales:
H_x: X/a≻X/b≻ . . . X/z ["≻" means "is more harmonic than"]

H_Y : Y/z≻ . . . Y/b≻Y/a
The *constraint alignment* is the pair of hierarchies:
*X/z≫ . . . ≫*X/b≫*X/a
*Y/a≫*Y/b≫ . . . ≫*Y/z

Combining the scales in (14) and (15) gives us the following constraint hierarchies:[9]

(17) *Onset sonority:* *Ons/Vowel≫*Ons/Liquid≫*Ons/Nasal≫*Ons/ Obstruent

(18) *Nucleus sonority:* *Nuc/Obstruent≫*Nuc/Nasal≫*Nuc/Liquid≫ *Nuc/Vowel

This pair of hierarchies, as others in Harmonic Alignment theory, has a special status in Con. Whereas normally, the rankings of constraints may freely vary from language to language (see §21.2.3.3), the relative ranking of the constraints above *with respect to each other* is universally fixed. They can be interspersed with other constraints, so they do not need to be adjacent in a specific language's hierarchy, but it can never be the case that *Nuc/Liquid dominates *Nuc/Nasal, for example, making syllabic liquids more marked than nasal ones.

This explains a well-established typological property of syllabification (Bell 1978): if less sonorous segments can be nuclei in a language, then the language must also allow all the more sonorous segments to be nuclei. Thus, as shown in (19), in some languages, only the most sonorous segments such as vowels may serve as syllable nuclei, whereas in others syllable nuclei can include vowels and liquids, and in still others—vowels, liquids, and nasals. No language allows nasal syllable nuclei without also admitting liquid and vocalic ones, all else being equal. The converse holds for syllable margins (Clements 1990, Steriade 1988, Hankamer and Aissen 1974).

(19) Sonority of syllable nuclei: a typology[10]

	Vowels	Liquids	Nasals	Obstruents
Spanish, Russian	✓	✗	✗	✗
Macedonian, Czech	✓	✓	✗	✗
English (unstressed syllables), Setswana	✓	✓	✓	✗
Berber, Central Carrier	✓	✓	✓	✓
Unattested	✗	✗	✓	✓

[9] Prince and Smolensky call their constraints *P/x and *M/x for "peak" (= nucleus) and "margin" (= onset), respectively, and their sonority scale includes more detail—which I abstract away from here.

[10] Language sources: Spanish (Harris 1983), Russian and Czech (Townsend and Janda 1996), Macedonian (Crosswhite 2001), English (Borowsky 1986), Setswana (Coetzee 2001), Berber (Prince and Smolensky 2004, Dell and Elmedlaoui 1985), Central Carrier (Walker 1979).

978–0–19–954400–4 21-Heine-C21-drv Heine-Narrog (Typeset by Spi, Chennai) 545 of 1025 June 30, 2009 16:19

Fixed rankings are designed to explain this typological observation. Any constraint that dominates *Nuc/Nasal, for example, will have to dominate *Nuc/Liquid and *Nuc/Vowel, since they are universally ranked below *Nuc/Nasal. A language with this ranking will tolerate vocalic, liquid, and nasal nuclei rather than violate the relevant constraint. For a concrete example, consider Standard American English. In English, a word-final two-consonant cluster is syllabified into a separate syllable if the second consonant is a sonorant, but not if it is an obstruent.

(20) English syllabification (Borowsky 1986, Levin 1985)

syllabic	liquids	syllabic	nasals	no syllabic obstruents	
peɪ.pɹ̩	'paper'	bʌ.ʔn̩	'button'	mɪks (*mɪ.ks̩)	'mix'
bʌ.kl̩	'buckle'	rɪ.ðm̩	'rhythm'	æsk (*a.sk̩)	'ask'

The conflict here is between creating a tautosyllabic consonant cluster, which violates the markedness constraint *Complex, and tolerating a consonantal nucleus, which violates one of the *Nuc/x constraints. In English, the solution is to tolerate sonorant consonantal nuclei but not obstruent ones, which suggests the ranking in (21). Syllabification patterns in English are thus non-uniform, with a cut-off point at nasals for minimum nucleus sonority.[11]

(21) Syllabification in English: sonorant but not obstruent consonants are parsed as nuclei

	*Nuc/Obs	***Complex**	*Nuc/Nas	*Nuc/Liq	*Nuc/Vowel
peɪ.pɹ̩~peɪpɹ		W		L	
bʌ.ʔn̩~*b*ʌʔn		W	L		
mɪks~mɪ.ks̩	W	L			

Compare the English pattern with that of Russian. Russian has only vocalic nuclei and freely tolerates margin clusters. The ranking of *Complex with respect to the *Nuc/x hierarchy in Russian must be as in (23):

(22) Russian syllabification: no syllabic consonants at all

metr	'meter'	dogm	'dogma Gen. Pl.'
so.fokl	'Sophocles'	fe.niks	'phoenix'

(23) Russian syllabification: tautosyllabic clusters are chosen over consonantal nuclei

	*Nuc/Obs	*Nuc/Nas	*Nuc/Liq	***Complex**	*Nuc/Vowel
metr~me.tr̩			W	L	
dogm~do.gm̩		W		L	
fe.niks~fe.ni.ks̩	W			L	

[11] An additional complication in English syllabification (in unstressed syllables) is that consonants may not by syllabic after a segment of greater sonority; thus, we get [fʌ.nl̩] "funnel" but [kɪln] "kiln". Relational constraints on sequences are also derived by schemata building on Harmonic Alignment; see Baertsch (2002), Gouskova (2004).

To complete this typology, consider the other rankings of *Complex with respect to the *Nuc/x hierarchy. If *Complex is ranked between *Nuc/Nasal and *Nuc/Liquid, the resulting grammar allows only liquid and vowel nuclei, whereas nasals and obstruents will be syllabified into clusters. This is what we find in a number of Slavic languages such as Czech and Macedonian. In Imdlawn Tashlhiyt Berber, on the other hand, any consonant may serve as a syllable nucleus, so sequences of consonants are syllabified into their own syllables rather than into margin clusters. The resulting typology is shown in (24):

(24) Factorial typology of *Nuc/x and *Complex

***Complex**≫*Nuc/O≫*Nuc/N≫ *Nuc/L≫*Nuc/V	any segment can be syllabic	Berber
*Nuc/O≫***Complex**≫*Nuc/N≫ *Nuc/L≫*Nuc/V	syllabic sonorants, but not obstruents	English
*Nuc/O≫*Nuc/N≫***Complex**≫ *Nuc/L≫*Nuc/V	syllabic approximants	Czech
*Nuc/O≫*Nuc/N≫*Nuc/L≫ ***Complex**≫*Nuc/V	only vowels can be syllabic	Russian

By now it should be apparent why the hierarchy of *Nuc/x constraints must be fixed. If *Nuc/x constraints could be reranked with respect to each other, then the theory would not make any predictions regarding typological implicational universals. If the ranking Nuc/Liq≫*Complex ≫*Nuc/Obs were possible, we would expect to see languages that have syllabic obstruents but not liquids. Such languages are unattested, so the possibility of such a ranking must be excluded.

The usual motivation for fixed rankings is that they reflect extragrammatical principles. The reason the *Nuc/x hierarchy is fixed is that it is based on the sonority scale, which reflects physical properties of sounds (such as intensity; see Parker 2008 for a recent overview). The strongest version of OT would only admit externally motivated universally fixed rankings, since fixed rankings are a kind of stipulation.[12] Indeed, there are plenty of proposals for universal constraint hierarchies that are not generated by Harmonic Alignment but are still based on phonetic and perceptual scales (Kirchner 1998, Flemming 1995, Steriade 2001, Kawahara 2006, de Lacy 2002*a*). There is a broad consensus in the literature on OT that many phonological markedness constraints are substantively grounded. Whether it is possible to reduce all phonological constraints to primitives, however, is a subject of ongoing work (Hayes 1999, Hayes et al. 2004, Smith 2002).

12 Another approach to hierarchies is to formulate constraints so that no matter how they are ranked, the universally most marked structures such as syllabic obstruents remain more marked than syllabic sonorants (Prince 1998, de Lacy 2004).

21.2.3.3 *Typology in OT*

As other generative theories, OT aims not only to delimit the range of cross-linguistic variation but also to derive universals. Both of these questions are addressed through a single mechanism: constraint re-ranking. OT offers a novel and strong hypothesis regarding cross-linguistic typology. According to the hypothesis, the range of cross-linguistic variation is determined by the number of constraint rankings that yield distinct sets of surface forms and mappings. The number of distinct rankings of constraints is the factorial[13] of the cardinality of Con, n!. Since factorials get quite large as n increases (e.g., 6! = 720, but 8!=40,320), it is essential to demonstrate that OT does not overgenerate distinct grammars. For most realistic constraint sets, there are far more constraint rankings than there are distinct outcomes, since many constraints do not conflict with each other, and others only conflict when dominated by other constraints. To take a simple example, consider a miniature model of Con below, which consists of three constraints discussed in the Tonkawa example (Max-V, NoGem, and Stressed=Heavy). This constraint set has 3!=6 permutations, but only three distinct outcomes: languages with no vowel deletion around stressed syllables (such as Spanish), languages with deletion blocked between identical consonants (Tonkawa), and languages with deletion regardless of consonantal context (Klamath, according to Bakovic 2005).

(25) A mini-Con and factorial typology

Rankings	sample mappings	Pattern and language
Max-V≫S=H≫NoGem NoGem≫Max-V≫S=H Max-V≫NoGem≫S=H	/pataka/→(pá.ta)ka /patata/→(pá.ta)ta	Vowel does not delete (Spanish)
NoGem≫S=H≫Max-V	/pataka/→(pát)ka /patata/→(pá.ta)ta	Vowel deletes except between identical Cs (Tonkawa)
S=H≫NoGem≫Max-V S=H≫Max-V≫NoGem	/pataka/→(pát)ka /patata/→(pát)ta	Vowel deletes regardless of context (Klamath)

There are only three distinct outcomes in this typology, even though there are six rankings. The reason is that Max-V and NoGem do not really interact with each other. If Max-V dominates Stressed=Heavy, vowel deletion is not an option, so it doesn't matter where NoGem is ranked. If Stressed=Heavy dominates Max-V, on the other hand, then all that matters is the relative ranking of Stressed=Heavy and NoGem. Depending on the ranking, deletion will either

[13] "Factorial" (n!) is the number of permutations of n elements, which is 1 if n = 0 and n*factorial(n− 1) if n > 0. The factorial is calculated by multiplying all numbers m: 0< m ≤ n. Thus, the factorial of 2 is 1*2=2, the factorial of 3 is 1*2*3=6, and so on.

be blocked or not, but two of the rankings amount to the same outcome. The more complex the constraint set, the richer the possibilities for such interactions, of course, and the number of possible distinct grammars cannot be predicted simply from the number of constraints. The nature of the constraints in question is crucial to working out the typology of possible phonological systems. For this reason, factorial typology is the primary means of testing proposed constraints.

The second typological concern is addressing universals: structures and patterns that either occur in all languages or in none. In OT, universals hold when a structure is allowed to surface under any ranking or is ruled out under any ranking. An example of the first type is the purported phonological universal that all languages have open CV syllables. Prince and Bruce Tesar (2004, ch. 6) show that this must be the case as long as (a) no language lacks CV sequences underlyingly (see next section on this), (b) there are constraints banning onsetless and closed syllables, and (c) there are no constraints that ban onsets or require that syllables be closed. A simple example of the second type is that no language has only nasalized vowels; the presence of nasal vowels implies oral vowels. One way to derive this is by ranking the constraint against nasal vowels universally over the constraint against oral vowels (McCarthy and Prince 1995); this still predicts a grammar that has no vowels, however, and so another approach would be to assume that Con has a constraint against nasal vowels but not one that bans all oral vowels (Gouskova 2003).

21.3 OT, THE LEXICON, AND THE INPUT

This section deals with the status of the lexicon and the input in OT, which is often a source of confusion for newcomers to the theory. As shown in §21.2.3.2, OT has no constraints that apply only to the input. Markedness constraints apply to outputs, and faithfulness constraints compare inputs and outputs. One of the motivations for this is to address the duplication problem: it is often the case that the same constraint apparently applies to derived and to underived (≈ underlying) sequences. We saw this in Tonkawa (recall (1)): there are no tautomorphemic geminates, and vowel deletion is not allowed to create new ones. An OT account explains both observations by assuming that NoGem rules out geminates at the surface level, regardless of their source. We know why Tonkawa phonology cannot create new geminates, but how does it rule out underived ones without ruling them out from the input? The answer in an OT account is that, even if hypothetical geminate inputs existed, they would not map faithfully in Tonkawa. Positing hypothetical inputs with geminates reflects an assumption known as *Richness of the Base* (Prince and Smolensky 2004): the input to a language's grammar is not subject to language-specific restrictions and may contain structures not found on the surface.

To understand Richness of the Base, it may help to distinguish between lexical entries and inputs to the grammar. OT is actually not tied to specific claims about the contents of the phonological lexicon. Much work in OT tacitly shares the SPE assumption that each morpheme has a unique underlying form that specifies its idiosyncratic features. Whether this is valid or not, lexical entries are not the same as the inputs that an OT grammar must be able to handle. A grammar describes (among other things) the speaker's knowledge of what surface forms are legal in the language, and an OT grammar derives legitimate surface forms by filtering out all illegitimate inputs. To explain why Tonkawa lacks underived geminates, we must therefore show that even if they were submitted to the grammar for evaluation, they would not map faithfully:

(26) Tonkawa and Richness of the Base: hypothetical geminate inputs map unfaithfully

/piccena-/	NoGem	Ident-Length
pi.cen~pic.cen	W	L

Importantly, this is not a claim about the underlying representations of words in Tonkawa. The actual lexicon of Tonkawa need not have morphemes with underlying geminates, but the grammar can handle inputs with geminates nonetheless. The duplication problem is addressed here by attributing the ill-formedness of all geminates, regardless of their source, to the same constraint: NoGem.

One last point about Richness of the Base is that it is often impossible to know exactly what the hypothetical inputs map to. In the case of Tonkawa, we have a clue that (26) is on the right track, since morphologically derived geminates shorten to singleton consonants. The lack of certainty is not a grave concern, however: the analytical goal is to show that unattested structures are in fact ruled out by the analysis, and if there isn't any evidence as to their fate, the analysis simply leaves it underdetermined.

21.4 Acquisition and learnability

The advent of OT revolutionized the study of phonological acquisition because OT can directly address Jakobson's (1941) observation that child speech is less marked than adult speech. OT's constraints capture this observation directly, assuming that

978–0–19–954400–4 21-Heine-C21-drv Heine-Narrog (Typeset by Spi, Chennai) 550 of 1025 June 30, 2009 16:19

children start out with an initial ranking in which markedness constraints dominate faithfulness constraints (see Gnanadesikan 2004 and other contributions to Kager et al. 2004). The initial ranking idea evokes the theory of Natural Phonology (Stampe 1973), whereby children start out with universal natural rules and learn to suppress some of them. Unlike SPE, though, Natural Phonology never became influential as a theory of rules, and SPE itself had little to say about acquisition or learnability since it had no mechanism for rule learning.

Research in OT has been accompanied by parallel work on learnability almost from the very beginning (Tesar and Smolensky 2000, Prince and Tesar 2004, Hayes 1999, Boersma and Hayes 2001). To learn an OT grammar is to arrive at a constraint ranking that produces the mappings and surface forms of the target grammar without overgenerating (Prince and Tesar 2004, Hayes 2004). In a realistic setting, this would require working out not only the ranking but also underlying representations (Merchant and Tesar to appear), structural ambiguity (Tesar 1998), and other problems that must be addressed in any theory of learnability in phonology. Because the problem of learnability is so complex, research on learnability must proceed incrementally by addressing these questions one at a time, and it is still an area of ongoing work.

21.5 Variation

Variation has been the subject of keen interest in recent work in phonological theory, both in OT and in related theories such as Harmonic Grammar. This is unsurprising, since OT characterizes in an explicit way what it means for two grammars to differ. If variation is seen as the coexistence of two grammars (or subgrammars) within an individual or a community, then it naturally suggests an intuitive approach to variation: variation exists when conflicting constraints are not conclusively ranked. This can be implemented formally with only slight modifications to the basic architecture outlined in §21.2. In this section, I describe just two such approaches; the reader is invited to consult Coetzee and Pater (to appear) for an excellent recent overview of others (both in OT and other frameworks).

An influential approach to variation in OT is Partially Ordered Grammars (Kiparsky 1993, Anttila 2002). This theory revises the notion of the language-specific constraint hierarchy $\mathcal{H}$ by under-determining the rankings of crucially conflicting constraints. At the point of selecting the optimum, a specific ranking of the hierarchy must be chosen, but it is chosen at random from several alternatives. To see how this works, consider again Tonkawa. Recall that in Tonkawa compound vowel deletion may delete word-final vowels even if this brings two identical consonants together. According to Hoijer (Hoijer 1946, 1949), the consonants shorten to a single consonant, but it appears that there is some optionality to the rule. The relevant facts from (2) are repeated below.

(27) Tonkawa (de)gemination

/taʔane-nisʔo:yta-/	taʔan(n)osʔo:ta-	'to stretch (e.g., a rope)'
/yakona-nacaka-/	yakon(n)acaka-	'to kill (him) with a blow of fist'
/yakexe-xakana-/	yakex(x)akana-	'to push (it) down hard'

Importantly, this vowel deletion is obligatory, but the choice between consonant deletion and a double consonant is optional. In Partially Ordered Grammars, the constraints against consonant deletion and geminates would be tied even though they conflict with each other. At the moment of utterance, the speaker has to choose between the two rankings. The hierarchy is as follows. The constraint requiring vowel deletion at the end of the first member of the compound is Final-C "a prosodic word ends in a consonant," and it is categorically ranked on top. The constraints against geminates (NoGem) and deletion (Max-C) are tied in the next stratum, so sometimes, NoGem will be violated (*yakex-xakana-*), and, other times, Max-C will be violated (*yakex-akana-*). Since word-medial vowel deletion is obligatory and always blocked by NoGem, the rest of the rankings must be fixed as shown in (30):

(28) Tonkawa optional degemination: geminated variant

	/yakexe-xakana-/	Final-C	Max-C	NoGem	Max-V
a.☞	yakex-xakana-			*	*
b.	yakex-akana-		*!		*
c.	yakexe-xakana-	*!			

(29) Tonkawa optional degemination: degeminated variant

	/yakexe-xakana-/	Final-C	NoGem	Max-C	Max-V
a.	yakex-xakana-		*!		*
b.☞	yakex-akana-			*	*
c.	yakexe-xakana-	*!			

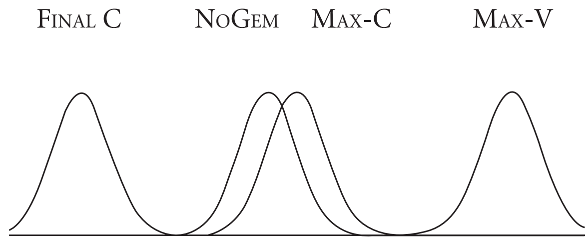
(30) Tonkawa hierarchy, in Partially Ordered Grammars

Final-C≫{NoGem, Max-C}≫Stressed=Heavy≫Max-V

Partially Ordered Grammars theory adds another dimension to capturing variation. Since the choice between rankings is random, it is assumed that the frequency of variants will depend on the number of rankings of the stratified hierarchy that produce those variants (see §21.2.3.3). In this analysis of Tonkawa, consonant deletion is predicted to occur half of the time, and geminate outputs should surface the other half of the time. In this way, Partially Ordered Grammars tries to account not only for the existence of variants but also for the frequency of their distribution.

Boersma and Hayes (2001) remark that Partially Ordered Grammars can only generate variant frequencies predicted by the constraint set, which is descriptively too weak. In reality, variant frequency may depend on factors other than the grammar, and it can also be strongly skewed toward one of the variants in a way that cannot be captured with a well-motivated constraint set. They add power to their model by redefining the hierarchy as probabilistic. Any given constraint has a probability range over which it is likely to be ranked. The greater the overlap between two constraints' ranges, the higher the likelihood of ranking reversal at utterance time. The hierarchy for Tonkawa would then look something like (31), and the degree of overlap between NoGem and Max-C could be varied to match the frequency of variants should it not be 50/50%.

(31) Tonkawa hierarchy, in Stochastic OT



Boersma and Hayes also propose a learning algorithm that, they argue, is not only capable of learning the target grammar but is also robust in the face of variation and can even reproduce frequencies of variants in the target grammar. With the development of approaches like this, phonological theory can now begin to broaden its empirical scope and address variation as an aspect of phonological competence.

21.6 Conclusion

Optimality Theory has revolutionized phonological theory more than any development since the SPE. It allowed phonologists to tackle problems such as conspiracies, typological differences and universals, phonological acquisition, learnability, and variation, all by introducing the profound claim that grammars consist of violable universal constraints. Still, there are many questions that have not been answered to everyone's satisfaction. Areas of ongoing work include phonological opacity, representations, the too-many-solutions problem, and the right approaches to exceptionality, lexical stratification, and issues at the interface of phonology and phonetics and syntax. At the same time, the theory is increasingly being tested using experimental and modeling methodologies from cognitive science (see the contributions to Coetzee et al. to appear). Thus, fifteen years after its arrival, it is still a vibrant theory with many directions for development.