Optimality Theory in Phonology

Alan Prince and Paul Smolensky

1. Architecture

GENERATIVE PHONOLOGY [*q.v.*] aims to construct a predictive theory of natural language sound systems, rooted in a finely-detailed account of the principles defining linguistic representations and the possible relations between them. Within these broad goals, OPTIMALITY THEORY (OT) develops in the context of specific empirical theses about the way phonological systems are organized. We identify three of them here.

(i) **The role of output targets**. Phonological representations may change from their lexical form to their surface or output forms: German /rad/, 'wheel – lexical form' is pronounced [rat]; Maori /inum/ 'drink - lexical form' is pronounced [i.nu], English /lik+d/ 'leak + past' is pronounced [likt], and so on. (Notation: we write '.' to demarcate syllables.) An important finding, announced in Kisseberth 1970 and since replicated in many areas of phonology and morphology, is that such changes are often conditioned by properties that hold of the surface or output forms in the language. In Maori, for example, no syllable is ever closed by a consonant: many lexical forms can be successfully parsed into open syllables without modification (e.g. /patu/ 'kill'), but others, like /inum/, must accommodate to the surface regularity by doing violence to lexical specification.

(ii) **Intrinsic relation between change and conditions of change.** Regular changes in representations take place under specifiable conditions; more subtly, the nature of the change is closely related to the nature of the conditions that provoke it. In Maori, for example, we see a change involving deletion under conditions that refer to a syllabic environment – but not just any syllabic environment is capable of forcing such deletion and not just any change may take place in that environment. The actual process must therefore have access not just to syllable structure *per se*, as a descriptive convenience, but to the properties of syllable structure which determine that consonant deletion, rather than (say) consonant insertion is the linguistically appropriate response. Theories of rule form in early generative phonology established little or no connection between the "structural description" or environment of a rule and the "structural change" it imposed.

(iii) **Universality and Difference**. Phonological grammars are strongly restricted both in absolute terms – some structures and patterns must always be present, others can never arise – and relatively, in that the presence of certain patterns implies the presence of others, or in that certain effects, if present, can only occur under certain conditions, cross-linguistically limited. Yet against this background of universality, languages may still differ considerably and arbitrarily – for example, it is not the case that every language disallows closed syllables, and it is not even the case that every language disallowing closed syllables will permit deletion to eliminate problematic structures that might give rise to them.

Optimality theory confronts these issues directly, designing its basic architecture to deal with them. The basic formal element of the theory is the *constraint*. Given the architecture of generative phonology, which distinguishes between underlying lexical forms and surface forms, and given the empirical finding (i) concerning the role of output targets, constraints must come in two basic species (Prince & Smolensky 1991, 1993):

(a) **Markedness constraints** evaluate *output* representations only, penalizing them for the presence of certain configurations.

A typical example is the constraint NoCoda, which forbids the occurrence in the output of syllables ending on a consonant (a syllable-final consonant is often called a 'coda' [*See* Syllable: Syllables in Phonology.]). Other examples include *Voi, which forbids the presence of voiced obstruents like [*b, d, z*] and so on, while being indifferent to the presence of their voiceless cognates [*p, t, s*]. Constraints may also be more specific in their demands; for example, there is good evidence for the existence of a constraint we can call *VTV, which has the effect of forbidding voiceless obstruents in intervocalic position: this constraint penalizes [opa] but passes on [oba] and [oa].

(b) **Faithfulness constraints** evaluate the relationship between input and output forms, demanding *exact replication* of the input along some specified structural dimension.

The constraint Max forbids all deletion: thus the input-output mapping *inum* → *i.nu* violates this constraint, but *patu*→ *pa.tu* does not. (Max, from McCarthy & Prince 1995, reformulates the effects of the constraint Parse in Prince & Smolensky 1993; its name alludes to the requirement that the input be *max*imally represented in the output). Along the same lines, the constraint Dep forbids insertion, and would be violated by e.g. *sta* → **e***sta*, since the output *e* is not present in the input. (The mnemonic, via McCarthy & Prince 1995, revising Prince & Smolensky's Fill, is that the output should *dep*end entirely on the input.) Featural identity is guaranteed by a family of constraints Ident(F), proposed in McCarthy & Prince 1995, such as Ident(Voi), which declare that any segment appearing in both input and output should have the same value of the feature F in both of its incarnations. For example, *opa*→*oba* violates Ident(Voi) in the change of /p/ to [b], but mappings such as *opa*→*opa*, or *opa* →*ota*, and even *opa* →*oa* all satisfy the constraint, the last one because the input segment *p* has no output avatar to be measured against.

If phonology is the computational link between the lexicon and phonetic form, then Markedness is the advocate of the phonetic interface, Faithfulness the agent of the lexical interface. The bipartite character of the constraint set is a direct projection of the computational role of the phonological module. Furthermore, it is only because the demands of the phonetic and lexical interfaces *conflict* that the grammar has a role for an active phonology which generates alternations.

Constraints often stand in direct disagreement with one another over the quality of a given input-output relation, as may easily observed even in the few examples already discussed. (To emphasize the Markedness or Faithfulness status of the cited constraints, we prefix them with M and F.)

• *inum*→ *i.nu.* satisfies M:NoCoda by virtue of output *i.nu.* but violates F:Max because *m* is lost from the input.

• *inum* → *i.num.* violates M:NoCoda but satisfies F:Max.

• *opa* → *oba* satisfies M:*VTV but violates M:*Voi, both by virtue of output *oba*, and violates F:Ident(Voi) as well, because p→b.

• *opa* → *opa* violates M:*VTV but satisfies both M:*Voi, both by virtue of output *opa*, and F:Ident(Voi) because p→p.

Such observations may be perspicuously displayed as in the *tableau* (1) where each data cell records the extent to which a mapping violates a constraint, one star per violation:

(1)

/inum/	M:NoCoda	F:Max
a. inum → i.nu.		*
b. inum → i.num.	*	

To arrive at OT, we invest the theory with a mechanism for actively employing such constraints to generate grammatical relationships. Any individual constraint offers a clear judgment, but the aggregate has many contending voices. How shall we find the concord of this discord?

Taking conflict to be inevitable, we drop the idea that the constraints in a grammar *must be obeyed*, must hold true of all licit mappings. Instead we regard the constraints as criteria of decision, which cooperate to select from among the various possible ways of realizing underlying forms. The input-output relation is thus construed as an optimization problem: given the range of alternative outputs for an input, and given the constraints, we do not ask "which output is the perfect match for the input?" but rather "which is best?". The primary function of constraints, then, is to compare alternatives: the reckoning of violations is just the first step toward making comparisons. In (1), for example, under this construal, the import of NoCoda is that (a) is better than (b), because (a) incurs fewer violations of the constraint than (b) does. The force of Max is the opposite: that (b) is better than (a).

To define 'best' in these circumstances, OT hypothesizes that constraints are *prioritized* with respect to each other on a language-specific basis. If a constraint A is prioritized above B, we will write A>>B and say that A is *ranked* above or *dominates* B. A ranking of the constraint set — a *constraint dominance hierarchy* — allows the entire set to evaluate alternatives. In (1), for example, if NoCoda is ranked above Max, then option (a) is deemed better than (b) over the entire constraint set. If, however, we have the opposite ranking, Max>>NoCoda, then (b) will be judged superior. Ranking the constraints into a single hierarchy puts all the dimensions of Markedness and Faithfulness on the same grammatical footing. The calculus of violation and comparison allows the demands of every constraint to be weighed on the same scale, despite considerable formal disparities among the dimensions along which linguistic structures vary.

Universal Grammar, constructed in accord with these ideas, will therefore contain a component *Con* which enumerates all its constraints, and a component *Gen* which defines for each input the set of outputs that it may be connected with.

Con tells us what the substantive constraints are, from which grammars are built. *Con* is not a grammar, but merely a set of often conflicting constraints. To arrive at a grammar, we impose a prioritization or ranking on the constraints in *Con*. It is a central hypothesis of Optimality Theory that a grammar ranks all the constraints in *Con* and that any ranking of *Con* is a grammar. (In this, OT picks up a line of thought due to Stampe 1973/79, in which a universal set of 'processes' was hypothesized to be fully present in the initial state, with grammars created by ordering and omission; OT disallows omission.)

Gen enumerates the range of alternative input-output pairs, from which the optimal or best one must be selected by a grammar. If grammar is optimization, and optimization is choice, then grammar must explicitly define, for each possible input, exactly what its possible alternative outputs are. (And, as well, to be precise, what *correspondence relation* relation the input has to each of the outputs, so that in comparing, for example, /inum/ and [i.nu], it is explicit that *m* is lost. See McCarthy & Prince 1995.) The collection of alternatives produced by *Gen* for a given input is called a *candidate set*. A second central hypothesis of Optimality Theory is that *Gen* is entirely universal and not susceptible at all to language-particular modification. Every language has the same range of choices, the same candidate set for each possible input: it is constraint-ranking alone that governs how the choices are made.

The third key component of Universal Grammar is a precise definition, which we may call *Eval*, that spells out what it means to be *optimal* with respect to a ranking of *Con*. As with the other components, *Eval* is taken to be entirely universal: there is no language-particular meddling with the meaning of optimality.

Although the notion of prioritization seems intuitively obvious, it is fraught with subtleties and care should be taken in using it. Let us approach it in terms of paired comparisons. Comparing two candidates, any given constraint can take one of three positions on the competition between them: it prefers one (by virtue of its having fewer violations); it prefers the other; or it judges them to be equivalent (by virtue of their having the same number of violations) and so fails to distinguish them in terms of the property being assessed. In the last case, the constraint simply does not bear on the choice between the two candidates; only the constraints that distinguish them can do that. Given a grammar, a complete ranking of *Con*, we say that candidate α is 'more harmonic than' or 'better than' candidate β, if α is preferred by the *highest-ranking constraint* that distinguishes α from β. A candidate ω is *optimal* if it is better than *every* distinct alternative candidate. ('Distinct' means that the constraint set can tell them apart, because at least one constraint distinguishes them.)

Even at this level of generality, it is possible to see that the theory succeeds in addressing the properties of phonological grammars that the discussion began with.

• **Markedness/ Faithfulness Interaction and Property (ii)**. Consider the grammar in which *all* faithfulness constraints are ranked above all markedness constraints. Then any candidate featuring a proscribed divergence from the input – for example, *inum*→*i.nu*, violating MAX – will lose out, somewhere in the Faithfulness subhierarchy, to the entirely faithful candidate *inum*→*i.num*, which satisfies all Faithfulness constraints perfectly. (We assume that there are no Faithfulness constraints concerning syllable structure; it is determined entirely by Markedness.) Consequently, if a lexical form is to undergo any kind of change in the phonology, some Markedness constraint must be ranked above, must 'dominate', a (relevant) Faithfulness constraint.

The immediate consequence is that any such change (the breach of Faithfulness) is intrinsically and indissolubly related to the conditions of change – Property (ii) above. The conditions of change emerge from the dominating constraints, at least one of which must be from the Markedness family. In the case at hand, the deletion is admitted only when **M**:NOCODA>>**F**:MAX. Violation of Faithfulness (MAX) is allowed, indeed required, but only to improve performance on the dominating Markedness constraint NOCODA. It is quite impossible to formulate a grammar that deletes under conditions, no matter how easily described, that have no connection to Markedness.

• **The Output targets** of Property (i), as promised, play a crucial role in determining the action of the phonology, because Markedness constraints evaluate output structures only. This result is strengthened by the key property of Faithfulness constraints, that they can only demand exact preservation of input characteristics. The M>>F configuration drives adaptation to output requirements, but there is analogous input-oriented mechanism that would invoke constraints on underlying configurations to provoke breaches of Faithfulness.

• **Universality and Difference.** The theory provides two sources for observational universality. First, and most obviously, various truths may be directly hard-wired into *Gen*, which lists linguistic primitives and defines their inalterable modes of combination.

The second, more interesting because unstipulated, comes from *Con* and its rankings: many candidates are simply never optimal in any grammar. As an illustrative, if miniaturized, example, consider the mapping *inum*→*i* in an OT nano-system where *Con* contains only NoCoda and Max. If *Gen* allows the omission of a segment from the input, it cannot be prevented (without stipulation) from allowing the omission of any number of segments. Under any reasonable *Gen*, the IO relation *inum*→*i* is going to be one of the candidates clamoring for optimal status. And it even has virtues: it satisfies NoCoda. But let us now compare it with the desired optimum *i.nu*.

(2)

/inum/	M:NoCoda	F:Max
a. inum → i.nu.		*
b. inum → i.		***

A moment's reflection will reveal that candidate (b) can never beat candidate (a) over this constraint set. Any decision between them must be made by Max; but Max prefers (a). Ranking will not help. Even though the input-output relation *inum*→*i* is an entirely legitimate competitor according to the laws of form encoded in *Gen*, it is not a possible optimum. The mapping it exemplifies *can never happen*.

This effect is called *harmonic bounding*: in the order of relative goodness imposed on the set of candidates by the constraints, no matter how they are ranked, candidate (a) is always better than (b) – it 'bounds' (b) – because (a) is better than or equal to (b) on every constraint in *Con*, and better on at least one. How important is harmonic bounding in more realistic settings? Samek-Lodovici & Prince (1999) note that in the numerous prosodic systems studied in Tesar (1999), involving 10 constraints, fully 3/4 of the four-syllable candidate forms are harmonically bounded; among five-syllable forms, the ratio is only slightly less at 5/8. A further effect worth noting in this example is that low priority – subordination in the ranking – is not equivalent to mere elimination. The constraint Max is subordinated to NoCoda, yet it is crucially Max that dismisses candidate (b). Whether or not a constraint is decisive in the selection process for any given candidate depends sensitively and precisely on the behavior of the constraints that dominate it.

The architecture of the theory places interlinguistic difference within the context of strong universality. It is not just the tools of analysis that are deemed universal, an empty gesture if no consequences or limitations follow. Substantive implications – desired or not – are logically inevitable as soon as commitments are made to the contents of *Con* and *Gen*. Any language-particular analysis requires specification of the relevant fragments of *Con* and *Gen*, which then inescapably dictate which grammars are universally possible, and which are not. Every analysis of a phenomenon in one language is necessarily a theory of that phenomenon in all languages.

2. Factorial Typology

The entire collection of rankings of *Con* determines what UG has to say about language. For a smaller set of constraints, the various rankings generate a 'factorial typology' of the predictions that those constraints make in interaction with each other. A factorial typology will not be the last word on a phenomenon if it doesn't include all relevant constraints; but it can still be analytically indispensable. As in the canonical sciences, where close study of simplified models has established itself in recent centuries, we will find that if the predictions and consequences are known under some degree of idealization, attention can be directed to their fate when idealization is modified or relaxed.

Let us focus first on the typology generated by just three syllable-structure-relevant constraints: M:NoCoda, F:Max, and F:Dep. Our approach will be informal. Suppose these constraints are challenged by an input that can give rise to a closed syllable as an optimum; let us chart the violation data for some of the key candidates:

(3) Sketch of Coda typology

/num/	M:NoCoda	F:Max	F:Dep	Remark on IO mapping
a. num→ .num.	*			faithful: closed syllable
b. num→ .nu.		*		deletion: open syllable only
c. num→ .nu.mA.			*	insertion: open syllables only

We assume for present purposes that *Gen* demands complete syllabification; that syllables contain a single vowel that may be flanked with consonants; and that *Gen* admits both deletion and insertion. For convenience, the inserted vowel is shown as *A*; in a fuller theory, its quality would be selected by Markedness constraints. We note here but do not explicitly list various harmonically-bounded candidates, such as those involving multiple deletion (*num→u*), multiple insertion (*num→A.nu.mA)*, pointless insertion (*num→A.num.*), pointless syllabification (*num→num.A*), and so on. The reader might wish to confirm their bounded status.

The high degree of symmetry in (3) makes it unusually easy to see how the typology plays out. It is a general fact of life that any constraint dispreferring an optimum in the face of its competitors must be subordinated. Here each candidate is dispreferred by just one constraint, and the other constraints jointly disprefer all of its competitors. Consequently, a given candidate will emerge as optimal when its antagonistic constraint is ranked beneath both of the others. Suppose for example we want (*cod*) to be the optimal output. Either of two rankings will produce this result:

(4) F:MAX >> F:DEP >> M:NOCODA
F:DEP >> F:MAX>> M:NOCODA

The ranking arguments are portrayed in the following tableau, where each row displays the fate of a comparison between the desired optimum and one of its competitors. The desired optimum is always indicated first in the pair $x \sim y$. For conciseness, the shared input is pulled out and written only in the upper left cell; recall, however, that a candidate is an entire input-output mapping. "W" indicates that a constraint prefers the desired optimum (or 'winner'); "L" that it prefers the desired suboptimum (or 'loser'), and blank indicates that the constraint fails to distinguish the competitors.

(5)

/num/ →	F:MAX	F:DEP	M:NOCODA	Remark on Comparison
a. num. ~ nu.	W		L	deletion worse than coda presence
b. num. ~ nu.mA		W	L	insertion worse than coda presence

Since the highest-ranking constraint that distinguishes the optimum from its competitor must in each case prefer the optimum, we deduce from (a) that F:MAX>>M:NOCODA and from (b) that F:DEP>>M:NOCODA. No ranking is required here among the faithfulness constraints. We can summarize the two resulting hierarchies in this way:

(6) {F:MAX, F:DEP}>> M:NOCODA.

By symmetry of the violation pattern (3), all systems will have the same shape: $\{C_1, C_2\}>>C_3$. For example, if a language has no codas at all, with problematic inputs handled by deletion, we have the following:

(7)

/num/ →	M:NOCODA	F:DEP	F:MAX	Remark on Comparison
a. nu. ~ num.	W		L	coda presence worse than deletion
b. nu. ~ nu.mA		W	L	insertion worse than deletion

As before, either ranking of the top two constraints will work, because the optimum, here *num→nu*, satisfies both of them and therefore cannot lose on them against any competitor. An argument for the crucial subordination of one constraint to another always involves a confrontation between a constraint that prefers the desired optimum (W) and another that prefers a competitor (L). No constraint preferring a suboptimal competitor can ever be allowed to be decide a comparison.

We therefore arrive at a full factorial typology of three possible systems:

1. No codas allowed. Deletion.
   {M:NoCoda, F:Dep} >> F:Max
2. No codas allowed. Insertion.
   {M:NoCoda, F:Max} >> F:Dep
3. Codas possible. No deletion or insertion allowed.
   {F:Max, F:Dep}>> M:NoCoda.

Even at this level of simplification, it is possible to notice interesting effects.

First, we see that a language may allow syllables to have codas, even though there is no markedness constraint favoring codas and, indeed, the only relevant markedness constraint actively proscribes them. Codas will appear when all routes of escape are blocked by sufficiently high-ranking Faithfulness constraints.

Second, no language can *demand* that syllables must have codas. Observe what happens when the constraint set is appropriately challenged by an input that doesn't provide good coda material:

(8)

/nu/ →	M:NoCoda	F:Dep	F:Max	Remark on Comparison
i) .nu. ~ .nuT.	W	W		nothing prefers suboptimum
ii) .nu. ~ Ø			W	nothing prefers suboptimum

The result is that an input like /nu/ must be preserved and syllabified *as is*, regardless of the ranking of the constraints. The alternatives are harmonically bounded – they have nothing going for them.

Third, even if codas are generally allowed in a language, they will still be avoided wherever that is possible with no breach of faithfulness. An input like /aka/ offers a choice between coda-ful *ak.a* and coda-less *a.ka.* Only the second of these can ever be optimal. Consider the fate of input /aka/ in the coda-admitting grammars (4) where Faithfulness is entirely dominant.

(9)

/aka/ →	F:Max	F:Dep	M:NoCoda	Remark on Comparison
a. ka ~ ak.a			W	nothing prefers suboptimum

This means that codas are, in essence, no more than tolerated under the Faithfulness-dominant regime: as soon as the opportunity arises to spurn them, they are abandoned. This is due to the emergent activity of the (thoroughly dominated!) NoCoda constraint.

Claims about the general patterns in inventories rest on a key assumption of OT: that the grammar is entirely defined by a ranking of *Con*. No other constraints or systems are entertained; in particular, there cannot be any as-yet-unarticulated subsystem with the responsibility of limiting the possible *inputs* to a grammar. With no limiting conditions, any input whatever is therefore possible, and each grammar must face the same, universal set of possible inputs. (This thesis is known as 'richness of the base'.) Potential input to the grammar cannot be construed as whatever happens to have ended up in the lexicon, as described by some language-particular set of lexical conditions [*See* Lexical Phonology: Morpheme Structure Constraints.]; no such conditions are recognized. In sharp contrast to the input set, the set of *outputs* can be strongly restricted by the grammar, as we have seen, and this is what shapes the observed inventory. To say that a language allows an element α in its inventory is to say that the grammar maps at least one of the universally possible inputs to α. To say that everything in the inventory has a certain property is to say that, no matter what the input is, the corresponding optimal output must have that property.

Applying these criteria to the preceding discussion discloses points where, for expositional purposes, logical steps have been omitted or abridged. To demonstrate that α is optimal, one must show that there is nothing in the candidate set that is better: so the entire candidate set must be addressed. To say that something holds of every output is to speak implicitly of every input. Such demonstrations may be found in the more thorough works in the literature.

To obtain a basic but realistic typology of syllable structure, it is necessary to add a constraint on onsets: ONSET, which requires that every syllable begin with a consonant. Taken in context with MAX and DEP, the analysis of its factorial typology mirrors that discussed above. (Calculate the fate of a suitably problematic input like /aka/ or even /a/.) The result is that three types of language are distinguished: those that require onsets, using deletion of a vowel to attain the result; those that require onsets, using epenthesis of a consonant; and those thattolerate onsetlessness when the input forces it, via dominant Faithfulness. (Further articulation of MAX and DEP is indicated as well: ONSET interacts with DEP-C, violated by consonant epenthesis; NOCODA interacts with DEP-V.)

Since ONSET and NOCODA are independent, the results of the two typologies may be combined: the result is a refined version of a well-supported universal typology promulgated by Jakobson (1962): in the syllabic inventory of a language, onsets may be required or not, but never banned; codas may be tolerated or not, but never required. Jakobson's statement is concerned with whether or not a certain structure appears in some word of the language; the present account focuses entirely on the input-output mapping, and the inventory is emergent from how the universal input set transforms under that mapping. Hence the refinements: a detailed account of how the outputs targets are achieved, along with predictions about how individual forms are actually parsed into syllables.

3. Distributions

Like syllabic systems, segmental inventories are limited in various characteristic ways. Coarsely, whole classes of segments are sure to be absent; more finely, relations like complementary distribution or environmental limitations on contrast are commonly encountered as well [*See* Phoneme.]. Here we sketch an approach to such patterns.

Imagine first a language, like many, in which only voiceless obstruents appear. Without limitations on the input, the grammar must contend with possible inputs containing voiced obstruents. Since no voiced obstruents surface, any that appear in an input must be treated unfaithfully; for concreteness, suppose they are mapped into their voiceless counterparts. A necessary part of the grammar will be the ranking *VOI>>IDENT(Voi).

(10) *VOI>>IDENT(Voi)

/da/ →	M:*VOI	F:IDENT(Voi)	Remark on Mapping
ta ~ da	W	L	voiced obs worse than IO voicing disparity

To truly ensure *da*→*ta*, all other relevant Faithfulness constraints would have to be put in place, ranked above IDENT(Voi); we gloss over this part of the analysis. Observe that if the ranking is reversed to F>>M, we'd be on our way to establishing *da*→*da*, leading to the presence of both voiced and voiceless obstruents in the output inventory. Are we sure that the M>>F system (10) nevers gains voiced obstruents from any other source – perhaps /ta/ itself? The mapping *ta*→*da,* from input /ta/, is harmonically bounded by *ta*→*ta*, which violates neither constraint. This means that we will be able to retain *ta*, no matter how the constraints are ranked.

Now imagine a language in which voiced and voiceless obstruents are in complementary distribution: suppose that voiced obstruents are found only intervocalically, where no voiceless obstruents are allowed. We must therefore establish two additional mappings. One voices underlying voiced obstruents intervocalically: *ata*→*ada*. The other preserves underlying *voiced* obstruents in the intervocalic environment, blocking the general devoicing process in just this case, so that *ada*→*ada*.

(11)

/ata/ →	M:*VTV	M:*VOI	F:IDENT(Voi)
ada ~ ata	W	L	L

As shown in the tableau, the only constraint preferring the desired optimum is *VTV; both of the others prefer its competitor, which avoids voiced obstruents and IO disparity entirely. Therefore, *VTV must dominate both. Minimally, we have *VTV>>*VOI. Accepting *VOI>>IDENT(Voi) from above, we have then a complete ranking on these constraints:

(12) M:*VTV >> M:*VOI >> F:IDENT(Voi)

What effect will this grammar have on input of the form /ada/? The danger is that the general devoicing process will swoop in, destroying the complementary distribution we seek. The following competition indicates that the situation is well in hand.

(13)

/ada/ →	M:*VTV	M:*VOI	F:IDENT(Voi)
ada ~ ata	W	L	W

The highest-ranking constraint that distinguishes the competitors is *VTV, and it prefers the desired optimum.

Complementary distribution has been achieved by controlling the input-output mapping system through ranking of Markedness and Faithfulness constraints. Possible inputs include all of /da, ta, ata, ada/, but the output includes only [ta, ada].

Though our discussion has looked at simple distributions, this general kind of analysis applies as well when morphophonemic evidence is available to establish underlying forms with certainty. The interaction of Markedness and Faithfulness constraints through ranking defines the class of input-output relations in the phonology. Whether the input is a specific lexical form, established by standard generative phonological criteria, or merely a member of the set of possible inputs not necessarily linked to an observed word of the language, the grammar functions to relate input and output by selecting the optimum from among the candidates supplied by *Gen*.

4. The subordination spectrum

The general empirical thesis (iii) of section 1 above, which notes the entanglement of universality and difference, may be considerably refined, when re-examined through the glasses provided by Optimality Theory.

The subordination spectrum. The phonological alternations and inventories that are observed cross-linguistically are shaped by the avoidance of marked elements and configurations. Suppose *m* is a marked type of configuration, defined as a structure violating some Markedness constraint M. If *Con* includes constraints antagonistic to M under various conditions – a virtual certainty – the nature of cross-linguistic variation with respect to *m* is this:

• **Absence.** In some languages, where M is sufficiently high-ranked with respect to conflicting constraints, M-violating structures *m* will be absent entirely.

• **Conditional Presence.** In other languages, where M is less high-ranked with respect to its antagonists, *m* will be limited to special contexts, where M's violation is forced by one of the constraints that crucially out-ranks it. Across other languages as M is ranked lower and lower, *m* will appear in more and more contexts as more conflicting constraints out-rank M..

This entire pattern is encapsulated in the simple statement that M is in *Con*.

The example from the previous section provides a simple illustration. The marked element *m* is a voiced obstruent; M is the constraint *VOI. In a language that ranks *VOI appropriately high — above both *VTV and IDENT(voi) — no voiced obstruents appear in the inventory, because of devoicing as in (10). There is no voicing contrast at all, and only the universally least-marked pole of obstruent voicing is admitted. In a language that ranks *VOI in an intermediate position — beneath *VTV but above IDENT(voi) — voiced obstruents appear only in a special context: intervocalic position, as in (11). Finally, in a language that ranks *VOI beneath IDENT(voi), voiced obstruents appear in all environments. And if the Markedness constraint *VTV is also sufficiently low-ranked — beneath IDENT(voi) — the marked configuration of an intervocalic voiceless obstruent can also appear, allowing full voicing contrast in all positions.

5. Issues and Extensions

Work in Optimality Theory has opened up new lines of attack on familiar problems, disclosed unasked questions lost amid the descriptive plethora of early generative phonology, revived traditional insights and stances that could not previously be made sense of, problematized issues considered solved, and produced a significant body of results which move learnability, computation, and formal analysis into a central position in the phonological research program. The key to these developments has been the commitment to strong universality within an architecture of constraint interaction. This has led to the existence of a core model sufficiently complex to be of empirical interest, but sufficiently narrow to allow for exact analysis of its structure and predictions. Here we offer a few pointers to research themes, with no pretense of completeness or canon-formation; see McCarthy 2002 for a comprehensive bibliography, and that volume as well as Kager 1999 for a survey of the field.

Con. Constraints fall into various classes, and show parallel articulation across the classes. McCarthy & Prince 1993b introduces a class of *Alignment* constraints, which position linguistic constituents with respect to each other, according to a rigid formula (the idea develops from the edge-based syntax-prosody mapping of Chen 1987 and Selkirk & Tong 1990). General faithfulness constraints can be restricted positionally to produce related, more specific constraints (Beckman 1997). Markedness constraints as well may show a dual pattern of positional restrictions (Zoll 1998). Correspondence relations have been posited between input and output (discussed above), between Base and Reduplicant (McCarthy & Prince 1993a), between a word and related word forms from which it is not directly derived (Benua 1998); in each case, the entire class of Faithfulness constraints is replicated along the dimension of correspondence. Correspondence between words leads to Output-Output Faithfulness, which reinvigorates traditional notions of paradigmatic dependencies (Kenstowicz 1997, Burzio 1999), challenging cyclic and stratal conceptions of such relations [*See* Phonological Derivations: Cycle.]. Construction of complex constraints from primitive scales is explored in Prince & Smolensky (1993) and de Lacy (2002); logical construction of complex constraints from simple constraints in Smolensky (1995) and Crowhurst & Hewitt (1977). Alderete (1999) expands the basic repertory of constraint types to include *Anti-faithfulness*, to explain paradigmatic alternations under Output-Output correspondence.

Factorial typology and the subordination spectrum. The range of weight effects on stress patterns between total quantity sensitivity and total quantity insensitivity [*See* Metrical Phonology.] is explained by ranking typology in Alber (1999), that between various types of quantity sensitivity in Rosenthall & van der Hulst (1999) and Morén (1999). Rhythmic patterns involving nonstandard overlapping constituents are shown to be well-behaved under ordinary alignment assumptions, and form the basis of a new typology of metrical systems in Hyde (2001). The typology of unbounded stress systems depends crucially on alignment in Baković (1998), but depends on Positional Markedness in Walker (1997). Factorial typology in the domain of obstruent voicing is carefully analyzed in Lombardi (1999), making use of Positional Faithfulness. In studying the range of effects due to the constraints on geminates and the Obligatory Contour Principle (role in OT: Myers 1997), Keer (1999) shows how Faithfulness constraints must be formulated to achieve the absolute nondefeasible restrictions on geminate patterning while admitting licit variations. Improving the match between factorial typology and attestation leads to a reassessment of the nature of the constraints controlling multiple footing in Kager (2002).

Prosodic Morphology. Prosodic and other phonological effects on morpheme shape and placement have been argued to follow from constraint domination. Templatic restrictions are then entirely emergent from the domination of Base-Reduplicant Faithfulness by ordinary Markedness constraints – the 'generalized template hypothesis' (McCarthy & Prince 1995; Itô, Kitagawa, & Mester 1996; Urbanczyk 1996; Alderete et al. 1999; Spaelti 1999).

Opacity. Phonological opacity, whereby morphophonemic processes are construed as dependent on representations neither underlying nor surface, supplies the life-blood and prime raison-d'être of rule-ordering theories [*See* Phonological Derivations: Rule Ordering in Phonology.], but for the most part lies outside the purview of core Optimality Theory. Extensions of the theory aiming to embrace opacity include McCarthy (1999ab, 2002b), wherein certain losing candidates are selected as bases for additional Faithfulness, and work by Smolensky and Goldrick (Goldrick 1999), which argues for preservation of unrealized structure in output forms. Kiparsky in privately circulated work argues that opacity should be understood in terms of lexical stratification of OT grammars, an architecture first postulated in McCarthy & Prince 1993a [*See* Lexical Phonology.]. A partly-related issue is the restriction of phonological patterns to lexical subclasses; Itô & Mester (1998) argue for lexical strata distinguished by differential placement of Faithfulness constraints, resulting in a hierarchy of ever more permissive domains [*See* Borrowing: Loanword Phonology.].

Modes of Resolution. Post-hoc narratives explaining, e.g., 'why the obstruent lost its voicing' tend to present themselves when needed and otherwise hide unassertively in the background. Factorial typology is less accommodating: a constraint once introduced into *Con* will have its way regardless of our prior intuitions. The study of factorial typology has therefore disclosed a number of previously unimagined ways that problems with Markedness constraints can in principle (if not in fact) be resolved. For example, while the markedness of obstruent voicing (*VOI) can be avoided by devoicing and even sonorization., there is no known attestation of the *deletion* of a voiced segment to resolve *VOI, even though factorial typology of the constraints discussed above cannot but predict it (Lombardi 2001). Such discoveries are grist for the explanatory mill of phonological theory in general, and OT in particular. Wilson (2001) proposes a modification of the optimization procedure that may lead to further insight into such cases; the fundamental idea is to restrict the scope of Markedness constraints so that they do not impose a relative goodness relation between every candidate, but only between a certain designated subset of them.

Learning, computation, and formal structure. Because of the logic of strict domination, and because OT constraints do compare all members of a candidate set with each other, it turns out to be remarkably straightforward and efficient to determine a workable ranking from a set of pairwise-comparisons, and to determine as well when no such ranking can exist (Constraint Demotion Algorithms: Tesar 1995, Tesar & Smolensky 1998). This leads immediately to the development of formal procedures for inducing aspects of phonology from data, given OT, through error-driven learning, and opens many avenues to a full-scale reckoning with the full complexities of grammar learning (Tesar 1999, 2000; Tesar & Smolensky 2000)[*See* Learnability]. This work is marked by its dependence on the intrinsic properties of OT, and its low level of paragrammatical learning-specific apparatus. The discrete constraint demotion algorithm has been extended in the continuous stochastic realm by Boersma (1997). Other computational studies deal with the generability and parsing efficiency of OT in terms of finite-state automata and the Chomsky hierarchy (Eisner 1997, Frank & Satta 1998, Karttunen 1998) [*See* Computational Phonology.]. The classic subset problem has received treatment in terms of the hypothesis that a learning bias exists in favor of ranking Markedness above Faithfulness constraints (Smolensky 1996, Prince & Tesar 1999, Hayes 1999). A considerable literature dealing with child acquisition phenomena has arisen, exploring this hypothesis. Harmonic bounding as described above has been generalized to cover cases where a *set* of candidates gangs up to prevent another candidate from becoming optimal under any ranking: this leads to a complete characterization of the circumstances under which irremediable suboptimality is found; strikingly, Constraint Demotion in its recursive form proves to be central to efficient reckoning of bounding (Samek-Lodovici & Prince 1999). The logic of entailment and inconsistency among pair-wise comparisons under OT has been shown to be closely related to the relevance logics RM and RM3, and their logical operation of 'fusion' provides the basis for sound, efficient manipulation and analysis of ranking arguments (Prince 2002). Core OT, with only Markedness and Faithfulness constraints, has been demonstrated to possess a very general mapping property, whereby structure-changing mappings must always result in a decrease in Markedness (Moreton 1999); this seemingly abstract property has immediate concrete consequences for the shape of allowed chain-shifts, for example disallowing swaps (X→Y together with Y→X in the same grammar) and more generally disallowing 'circular' shifts. In consort with assumptions about the nature of complex constraints (Smolensky's 1995 'local conjunction'), a predictive and empirically-justified typology of possible chain-shifts emerges (Moreton & Smolensky 2002).

References

ROA= Rutgers Optimality Archive, http://roa.rutgers.edu

Alber, Birgit. 1999. Quantity Sensitivity as the result of constraint interaction. ROA-310.

Alderete, John. 1999. *Morphologically governed accent in Optimality Theory*. New York: Routledge. ROA-309.

Alderete, John, Jill Beckman, Laura Benua, Amalia Gnanadesikan, and John McCarthy. 1999. Reduplication with fixed segmentism. *Linguistic Inquiry* 30.327-364. ROA-226.

Baković, Eric. 1998. Unbounded stress and factorial typology. In *RuLing Papers* 1. New Brunswick: Rutgers University. ROA-244.

Beckman, Jill. 1997. *Positional Faithfulness*. Ph. D. dissertation. Amherst: University of Massachusetts. ROA-234.

Benua, Laura. 1998. *Transderivational Identity: Phonological Relations Between Words.* Ph.D Dissertation. Amherst: University of Massachusetts. ROA-259.

Boersma, Paul. 1997. How we learn variation, optionality, and probability. *Proceedings of the Institute ofPhonetic Sciences of the University of Amsterdam* 21.43–58. ROA-221.

Burzio, Luigi. 1999. Surface-to-Surface Morphology: when your Representations turn into Constraints. ROA-341.

Chen, Matthew. 1987. The syntax of Xiamen Tone Sandhi. *Phonology* 4.109-150.

Crowhurst, Megan and Mark Hewitt.1997. Boolean operations and constraint interactions in Optimality Theory. ROA-229.

de Lacy, Paul. 2002. *The Formal Expression of Markedness.* Amherst: Ph.D Dissertation. Amherst: University of Massachusetts. ROA-XXX.

Eisner, Jason. 1997. Efficient generation in Primitive Optimality Theory. *Proceedings of the 35th Annual Meeting of the ACL*. Madrid. ROA-206.

Frank, Robert and Giorgio Satta. 1997. Optimality Theory and the generative complexity of constraint violability. *Computational Linguistics* 24.307-315. ROA-228.

Goldrick, Matt. 1999. Turbid representations and the unity of opacity. In *Proceedings of the 30th Annual Meeting of the Northeastern Lingusitics Society*, edited by Masako Hirotani, pp. 231-246. Amherst, Mass.: GLSA. ROA-368.

Hayes, Bruce. 1999/to appear. Phonological Acquisition in Optimality Theory: the Early Stages. To appear in *Fixing Priorities: Constraints in Phonological Acquisition*, edited by René Kager, Joe Pater, and Wim Zonneveld. New York: Cambridge University Press. ROA-327.

Hyde, Brett. 2001. *Metrical and prosodic structure in Optimality Theory*. Ph. D. Dissertation. New Brunswick: Rutgers University. ROA-476.

Itô, Junko and Armin Mester. 1998. The phonological lexicon. In *The Handbook of Japanese Linguistics*, edited by Natsuko Tsujimura, pp. 62-100. New York: Blackwell. ROA-256.

Itô, Junko, Yoshihisa Kitagawa, and Armin Mester. 1996. Prosodic Faithfulness and Correspondence: Evidence from a Japanese Argot. *Journal of East Asian Linguistics* 5.217-294. ROA-146.

Jakobson, Roman. 1962. *Selected Writings 1: Phonological Studies*. s'-Gravenhage: Mouton.

Kager, René. 2002. Rhythmic Directionality by Positional Licensing. ROA-514.

Kager, René. 1999. *Optimality Theory*. New York: Cambridge University Press.

Karttunen, Lauri. The proper treatment of optimality in computational phonology. In *Finite State Methods in Natural Language Processing*, pp. 1-12. Ankara. ROA-258.

Keer, Edward.. 1999. *Geminates, OCP, and the nature of CON*. Ph.D. dissertation. New Brunswick: Rutgers University. ROA-350.

Kenstowicz, Michael. 1997. Uniform Exponence: Exemplification and Extension. ROA-218.

Kisseberth, Charles. 1970. On the functional unity of phonological rules. *Linguistic Inquiry* 1.291-306.

Lombardi, Linda. 1999. Positional faithfulness and voicing assimilation in Optimality Theory. *Natural Language & Linguistic Theory* 17.267-302.

Lombardi, Linda. 2001. Why place and voice are different: Constraint-specific alternations in Optimality Theory. In *Segmental Phonology in Optimality Theory*, edited by Linda Lombardi. New York: Cambridge University Press. ROA-105.

McCarthy, John. 1999a. Sympathy and phonological opacity. *Phonology* 16.331-399. ROA-252.

McCarthy, John. 1999b. Sympathy, cumulativity, and the Duke-of-York gambit. ROA-315.

McCarthy, John. 2002a. *A thematic guide to Optimality Theory*. New York: Oxford University Press.

McCarthy, John. 2002b. Comparative Markedness. ROA-489.

McCarthy, John J., and Alan Prince. 1993a. *Prosodic Morphology: Constraint Interaction and Satisfaction*. Rutgers Center for Cognitive Science Report 3. ROA-482.

McCarthy, John J., and Alan Prince. 1993b. Generalized Alignment. In *Yearbook of Morphology*, edited by Geert Booij and Jaap van der Marle, pp. 79-153. Dordrecht: Kluwer. ROA-7.

McCarthy, John and Alan Prince.1995. Faithfulness and Reduplicative Identity. In *Papers in Optimality Theory*, edited by Jill Beckman, Laura Walsh Dickey, and Suzanne Urbanczyk. Amherst: GLSA. ROA-60.

Morén, Bruce. 1999. *Distinctiveness, coercion, and weight.* Ph.D. dissertation. College Park: University of Maryland. ROA-336.

Moreton, Elliott. 1999. Non-computable functions in Optimality Theory. ROA-364.

Moreton, Elliott and Paul Smolensky. 2002. Typological consequences of local constraint conjunction.To appear in *WCCFL* 21. ROA-525.

Myers, Scott. 1997. OCP effects in Optimality Theory. *Natural Language & Linguistic Theory* 15.847-892. ROA-6.

Prince, Alan and Paul Smolensky. 1993. *Optimality Theory: Constraint Interaction in Generative Grammar.* RuCCS-TR-2, Rutgers Center for Cognitive Science. New Brunswick: Rutgers University. ROA-XYZ.

Prince, Alan and Bruce Tesar. 1999/ to appear.. Learning Phonotactic Distributions. To appear in *Fixing Priorities: Constraints in Phonological Acquisition*, edited by René Kager, Joe Pater, and Wim Zonneveld. New York: Cambridge University Press. ROA-353.

Prince, Alan. 2002. Entailed Ranking Arguments. ROA-500.

Rosenthall, Sam and Harry van der Hulst. 1999. Weight-by-position by position. *Natural Language & Linguistic Theory* 17.499-540.

Samek-Lodovici, Vieri and Alan Prince. 1999. Optima. RuCCS-TR-57. Rutgers Center for Cognitive Science. New Brunswick: Rutgers University. ROA-363.

Selkirk, Elisabeth. The prosodic structure of function words. In *Papers in Optimality Theory*, edited by Jill Beckman, Laura Walsh Dickey, and Suzanne Urbanczyk, pp. 439-70. Amherst: GLSA.

Selkirk, Elisabeth and Tong Shen. 1990. Prosodic domains in Shanghai Chinese. In *The Phonology-Syntax Connection*, edited by Sharon Inkelas and Draga Zec, pp. 313-37. Chicago: U. Chicago Press.

Smolensky, Paul. 1995. On the structure of the constraint component Con of UG. ROA-86.

Smolensky, Paul. 1996. The initial state and 'Richness of the Base' in Optimality Theory. ROA-154.

Spaelti, Philip. *Dimensions of variation in multi-pattern reduplication.* Ph.D. dissertation. Santa Cruz: University of California. ROA-311.

Stampe, David. 1973/79. *A dissertation in Natural Phonology.* Ph.D. Dissertation. New York: Garland Press.

Tesar, Bruce. 1995. *Computational Optimality Theory*. Ph. D. Dissertation. Boulder: University of Colorado. ROA-90.

Tesar, Bruce. 1999. Robust Interpretive Parsing in Metrical Stress Theory. *WCCFL* 17.625-639. ROA-262.

Tesar, Bruce. 2000. Using inconsistency detection to overcome structural ambiguity in language learning. RuCCS TR-58. ROA-426.

Tesar, Bruce, and Smolensky, Paul. 1998. Learnability in Optimality Theory. *Linguistic Inquiry* 29.229-268.

Tesar, Bruce B., and Smolensky, Paul. 2000. *Learnability in Optimality Theory*. Cambridge, Mass.: MIT Press.

Urbanczyk, Suzanne. 1996. *Patterns of reduplication in Lushootseed.* Ph. D. Dissertation. Amherst: University of Massachusetts.

Walker, Rachel. 1997. Mongolian Stress, Licensing, and Factorial Typology. ROA-172.

Wilson, Colin. 2001. Consonant cluster neutralization and targeted constraints. *Phonology* 18.147-197.

Zoll, Cheryl. 1998. Positional asymmetries and licensing. ROA-282.