The Universal Constraint Set: Convention not Fact*

T. Mark Ellison

All languages make the same phonological generalisations. This is the remarkable claim of Optimality Theory (OT).

In early generative phonology (Chomsky & Halle 1968), phonological generalisations were expressed by ordered rewrite rules. Each language, however, required its own set of rules as well as its own ordering. Later, underspecification phonology (Archangeli & Pulleyblank 1989, 1994) emphasised default rules. Universal tendencies in the rules were apparent, but characterising all languages with a single set of rules remained an unreachable dream.

In OT, phonological generalisations are expressed as ranked defeasible constraints. Ranking provides so many distinct but plausible grammars that it seems feasible that a universal set of phonological generalisations could account for the diversity of phonological systems.

The question we face is no longer whether the assumption of such a universal set is theoretically tenable, but whether it is justifiable.

There are two senses in which such an assumption could be justified: either as a fact or as a convention. If a fact, it claims that all language users objectively instantiate the same set of generalisations. If a convention, it encourages phonologists to describe languages using an agreed but arbitrary system of generalisations. In this interpretation, the universal constraint set is as arbitrary, but as useful, as the international phonetic alphabet (IPA).

This chapter examines seven kinds of argument for one or other status of the universality of phonological constraints. These are the arguments from empirical evidence (section 2), restrictiveness (section 3), simplicity (section 4), universal markedness, acquisition (both section 5), learnability (section 6), and convention (section 7). Close examination finds all but the last of these arguments to be wanting.

The conclusion that remains is that universality, like the IPA, makes a better convention than fact. It should be used rather than believed.

1 Optimality Theory and Universals

Optimality Theory (Prince & Smolensky 1993, for an introduction see Archangeli & Langendoen 1997) is first and most frequently applied to phonology[1]. In this

*I would like to thank the following people for their comments on an earlier version of this paper: Ash Asudeh, Paul Boersma, Patricia Cabredo Hofherr, Frank Keller, Alice Turk, Markus Walther, and one anonymous reviewer. I would also like to thank the editors for their patience and interest.

[1]The formalism has also been applied to morphology (e.g. Benua 1995, Golston & Wiese 1995, Orgun 1994, Russell 1995) and syntax (Dickey 1995, Grimshaw & Samek-Lodovici 1995,

domain, the theory defines a metalanguage for stating generalisations about phonological sequences and representations, and at the same time it determines how these generalisations interact when combined to form complex analyses. Although the individual concepts of OT are presaged in earlier literature, its combination of sweeping generalisations with a simple mechanism of combination has proved very popular in the phonological community.

Since the scientific study of phonology began, its practitioners have intuited many powerful generalisations, but exceptions have plagued attempts to give these generalisations a precise expression. Optimality theory offers a mechanism for protecting generalisations from the pernicious effects of exceptions: all exceptions to any constraint are either lexically required, or achieved by a conspiracy of more highly-valued constraints. While the lexical exception was a part of earlier phonological theories, they lacked principled mechanisms for capturing patterned exceptions.

Other chapters in this book have introduced and exemplified the basic concepts of Optimality Theory, so there is no need to present a detailed account of OT here. Rather, I propose to highlight those aspects of OT which will play a role in the material found later in the chapter.

1.1 Three Optimality Theories

The basic components of OT are: a lexicon which can provide input candidate sets, a ranked set of violable constraints, and an evaluation function which eliminates non-optimal candidates. These components, on their own, define a pure theory of constraint interaction, which we may refer to as *Pure Optimality Theory* (OT_0). OT_0 does not include any assumptions about what can be in the lexicon, how the candidate sets are generated, or what the constraints are. It only stipulates the generation mechanism.

Supplementing this theory with two further assumptions defines what I will call *Standard* OT ($\mathrm{OT}_{\mathrm{P\&S}}$), the theory proposed by Prince and Smolensky (1993). These additional assumptions are: **Gen** and **Univ**. The first of these assumptions (1) concerns the relation between the lexicon and candidate sets.

(1) The candidate sets for each utterance are generated from lexical representations by a universal function **Gen**.

The second assumption, **Univ**, will be discussed in section 1.2.

The initial statement of OT was very like Declarative Phonology (Bird 1990, 1995, Bird & Ellison 1994, Scobbie 1991, 1997, Scobbie et al. 1995) in its monostratal formulation: constraints acted only on surface forms, combining to eliminate all but the correct forms from those offered by **Gen**.

More recent work (McCarthy & Prince 1995, McCarthy 1996) has seen a shift towards incorporating a second level of *phonological* representation, usually identified with the lexical input to **Gen**. Phonological derivation therefore includes input and output levels of representation[2], and constraints control the relationship between these.

Note that this two-level approach is reminiscent to the finite-state transducer models of morphology and phonology (Koskenniemi 1983, Antworth 1990), an observation made by Orgun (1995).

Legendre et al. 1993, Sells et al. 1994, Speas 1995, Woolford 1995).

[2]Prince and Smolensky (1993:192) do presage the two-level approach.

In this two-level theory of OT ($OT_{2\text{-L}}$), constraints on the phonological output, the so-called *structure* constraints, supplement constraints matching lexical forms to surface forms, the *faithfulness* constraints[3]. The ranking of phonological structure and faithfulness constraints determines the compromise made between the demands of the lexical input and the pressure for unmarked surface forms.

1.2 The universal constraint set

The Optimality Theory of Prince & Smolensky (1993) assumes two universal components beyond the basic mechanism of constraint ranking. The first of these, **Gen**, creates candidate surface forms from lexical entries. The second defines the set of constraints, common to all languages. Variation between languages is accomplished not by having different constraints, but by modifying the priority rankings between them. In Prince and Smolensky's words (1993:5): *'constraints are essentially universal and of very general formulation'*. For ease of reference, **Univ** will denote this assumption of universality.

Univ lends itself to two distinct interpretations. According to the the stronger of these, it states a fact about the mental reality of language users (2). This strong assumption will be denoted **Univ-Fact**.

(2) **Univ-Fact**: There is (at least) one hierarchy of constraints objectively present in the mind of each language user. Furthermore, the same constraint set is used in each hierarchy of each and every user.

This strong form of **Univ** is implicit in much OT work, including the original technical report. As a typical example, Prince & Smolensky (1993:5) refer to the constraint hierarchy as a cause: 'interlinguistic differences arise from the permutations of constraint-ranking' — they do not arise from differences in the constraint set.

In a similar vein, Smolensky equates language acquisition with the manipulation of constraint rankings. 'In Optimality Theory, learning a target adult language requires a child to determine the relative rankings of universal constraints' (1996:17). Here, the child is assumed to have a mentally objective constraint hierarchy replete with universal constraints.

Archangeli (1997) also makes this assumption of universality a cornerstone of her account of Optimality Theory.

> CON, as a universal set of constraints, is posited to be part of our innate knowledge of language. What this means is that every language makes use of the same set of constraints. ... This is the formal means by which *universals* are encoded (p15).

In this chapter, I offer an alternative interpretation of **Univ**, which takes the uniformity of constraint description to be a methodological desideratum, rather than a statement of fact (3).

(3) **Univ-Conv**: Languages should be analysed (as much as possible) using a constraint set common to the community of phonologists.

[3] McCarthy & Prince 1995 also introduce the notion of constraints controlling correspondences between surface forms.

Some of the arguments presented in this paper contrast **Univ**, as either **Univ-Fact** or **Univ-Conv**, with the lack of this assumption. For ease of reference, this lack will be given the name **NoUniv**.

(4) **NoUniv**: Languages may or may not use the same constraints.

Like the original statement of **Univ**, **NoUniv** is ambiguous, referring to a lack of uniformity among either mentally real constraints, or the purely descriptive constraints of linguistic analyses. The context will serve to distinguish the senses, when the distinction is relevant.

Now that we have a precise notion of constraint universality to work with, we can proceed to the question of whether linguistic evidence could ever empirically show that objective constraints are universal, i.e. that **Univ-Fact** is true.

2 Empirical Evidence

The argument for a universal constraint set from empirical evidence is one I have never seen put forward, but it is certainly imaginable, and so, for the sake of completeness, takes its place here.

Many kinds of empirical evidence are imaginable, but few are found. We could, in a flight of fancy, imagine autopsies revealing neurons carefully inscribed with the names of their corresponding phonological or syntactic constraints. In reality, however, all empirical evidence for linguistic generalisations in the mind is indirect. We have access to: surface forms, variation in surface forms, meanings, and the results of elicitation. A generous interpretation of this evidence would claim that it suffices to identify both lexical candidate sets, and the corresponding optimal forms. Supposing this evidence were available, an empirical argument for a universal set of constraints might develop as follows.

(A) The Argument from Empirical Evidence

(A-1) Empirical evidence about the selection of optimal candidates from lexical candidate sets is collected for many languages.

(A-2) In each language **L**, the empirical evidence forces us to conclude that users employ a particular constraint hierarchy $\mathbf{H_L}$.

(A-3) All of these hierarchies $\mathbf{H_L}$ use the same constraints.

$\therefore$ All languages use the same constraints.

The weak assumption in this argument, without which it cannot succeed, is (A-2). We show below that for any constraint hierarchy, there is another which uses a different set of constraints but always selects the same candidates as optimal. Thus no amount of data can force us to conclude that a particular language uses a given constraint set: there is always an equally well-supported alternative.

2.1 Constraint addition

The basis for the counterargument is an operation for combining two constraints, an operation we can call *addition*. The addition **C+D** of two constraints **C** and **D** designates a distinct third constraint which assigns to each candidate the sum of the number of violations assigned by constraints **C** and **D**.

To illustrate addition, table 1 shows the evaluations assigned to various phoneme sequences, Portuguese words in this case, by the two well-known constraints **Ons**, requiring onsets, and **NoCoda**, prohibiting codas,and by their sum **Ons+NoCoda**. There is, of course, nothing special in the choice of the two constraints for this example. Any other two constraints would have sufficed equally.

	Ons	**NoCoda**	**Ons+NoCoda**
/pɐ.rɐ/	.	.	.
/ɐ.mo.rɨʃ/	*	*	**
/tur.ni.a.dɐʃ/	*	**	***
/al.ku.ɔl/	**	**	****

Table 1: The evaluation of candidates under summed constraints. The full stop is used to mark the absence of violations.

The word /alkuɔl/, orthographically <**alcool**>, *'alcohol'*, has two onsetless syllables and two codas, and so engenders two violations each to **Ons** and **NoCoda**. Consequently, it incurs four violations of the sum constraint **Ons+NoCoda**.

I should emphasise here that summed constraints, such as **Ons+NoCoda**, are independent, singleton constraints. They bear no relation to their component constraints, except the mathematical relationship in the number of exceptions.

2.2 Two equivalent hierarchies

Now consider the action of the two two-constraint hierarchies **Ons≫NoCoda** and **Ons≫Ons+NoCoda**. Table 2 shows these two hierarchies selecting among some candidate syllabifications of /subʃtituɐ/, *'substitute (3s subj)'*.

		Ons	**NoCoda**	**Ons**	**Ons+NoCoda**
☞	/su.bʃ.ti.tu.ɐ/	*	*	*	**
	/sub.ʃ.ti.tu.ɐ/	**	**	**	****
	/sub.ʃ.tit.u.ɐ/	***	***	***	******
	/su.bʃ.tit.u.ɐ/	**	**	**	****

Table 2: Equivalence of hierarchy with summed constraints.

The optimal candidate from the two hierarchies is the same, /su.bʃ.ti.tu.ɐ/. This is not a coincidence. Two hierarchies **C≫D** and **C≫C+D** will select the same optimal candidate whenever applied to the same candidate set. In both cases, the higher-ranked constraint elects candidates optimal to it, and the lower-ranked constraint need only choose between these.

In the first hierarchy, this means that of the optimal candidates according to **C**, the candidate(s) with the least violations to **D** will be regarded as optimal. In the second hierarchy, once again **C** dominates, and so of the candidates optimal according to **C** the candidates showing the least violations of **C+D** will be optimal to the hierarchy. But all candidates optimal in **C** will have the same evaluation for **C**, and thus the only differences in **C+D**'s evaluation of these candidates is provided by differences in **D**. Consequently, of the candidates optimal in **C**, those optimal in **D** will also be optimal in **C+D**. Thus precisely the same candidates incur minimal violations according to these two constraints. Therefore these two small hierarchies select the same optimal candidates.

2.3 Constructing distinct but equivalent hierarchies

Given any hierarchy with more than one constraint, we can construct a distinct, but functionally equivalent, second hierarchy by the simple expedient of replacing its second-ranked constraint by the sum of the second-ranked constraint and the first-ranked. For example, if the two highest ranked constraints in the first hierarchy were **Ons** and **NoCoda** in that order, then replace **NoCoda** with **Ons+NoCoda**, keeping all other constraints the same, to make a new hierarchy.

This new hierarchy has a different constraint set from the first; **NoCoda** is missing from the second constraint set. But as we have seen, the combined selective action of the first two constraints in both hierarchies is the same. As all subsequent constraints are identical, the action of the two hierarchies as a whole is identical. The two hierarchies can be regarded as notational variants for the same function. As the constraints in the two hierarchies are different, they cannot both accord with a putative universal candidate set.

Empirical evidence cannot ever distinguish between two functionally equivalent hierarchies. Consequently, empirical evidence alone can never identify a unique constraint set for a given language. The evidence which supports the putative universal constraint set in a language also always supports alternatives using different constraints.

It might be argued that **C**$\gg$**D** and **C**$\gg$**C+D** are uninteresting notational variants, lacking distinctive linguistic value. This is not the case, for precisely the reason that is important to this discussion. Reversing the rankings of these two hierarchies results in hierarchies that make different decisions on certain candidate sets.

For example, suppose **C** is **Ons** and **D** is **NoCoda**. The hierarchies **Ons**$\gg$**NoCoda** and **Ons**$\gg$**Ons+NoCoda** always select the same optimum from a candidate set. If the rankings are reversed, however, this is not the case. The candidate /kal/ violates **NoCoda** once, while /a.ka.la/ violates it not at all. However, /a.ka.la/ violates **Ons** once. Both candidates violate **Ons+NoCoda** the same number of times. So in the reversed ranking **Ons+NoCoda**$\gg$**Ons**, it is the candidate best satisfying **Ons** which is optimal. The corresponding ranking **NoCoda**$\gg$**Ons** prefers /akala/ as it offers no violations to the higher-ranked constraint **NoCoda**. These comparisons are tabled in tableau 3.

So while the two hierarchies offer the same weak generative capacity when ordered in these hierarchies, reversing the ordering results in different predictions. While the two hierarchies are notational variants, the differences in notation are linguistically important.

		NoCoda	Ons		Ons+NoCoda	Ons
/kal/		*!	.	☞	*	.
/a.ka.la/	☞	.	*		*	*!

Table 3: Parallel tableaux showing the different selective power of **NoCoda**$\gg$**Ons** and **Ons+NoCoda**$\gg$**Ons**.

2.4 Notational variants

These notational variants pose a serious problem for an objective interpretation of **Univ**. They mean that the hypothesis not be proven empirically. Furthermore, it fails to meet a primary criterion for psychological reality. Harman (1980:21) states of a true theory, namely one that is in accordance with the empirical evidence, that aspects *'not shared by its notational variants are not taken to have psychological reality.'* Coherence with a putative universal constraint set is not a property shared by all notational variants of any OT analysis of an individual language. The constraint set cannot, therefore, be ascribed psychological reality.

The conclusion, therefore, is that if we can analyse linguistic data using one constraint hierarchy, we can always use another hierarchy with a different constraint set to do the same job. The assumption (A-2) of the argument from empirical evidence always fails. Consequently, argument (A) can provide no support for **Univ-Fact**. **Univ-Fact** cannot be proven empirically.

Furthermore, by not being independent of notational variance, i.e. non-empirical variance, in language analyses, the universal constraint set fails a major criterion for psychological reality.

It might be argued, however, that the impetus for **Univ-Fact** is not simply empirical but indirect. The next four sections consider indirect arguments for **Univ**.

3 Restrictiveness

The second argument for a universal constraint set relies on the frequently cited desideratum of restrictive linguistic hypotheses. Smolensky (1996:3) includes restrictiveness among the advantages of encapsulating systematic cross-linguistic variation within constraint-ranking.

In much linguistic literature, including the article just cited, it is unclear whether restrictiveness refers to limitations on structure underspecified by the linguistic evidence, or whether it refers to predictive limitations of what might be observed. However, some works do emphasis the importance of predictive or empirical restrictiveness in allowing hypotheses to be tested. Chomsky (1978:9) identifies this, and the consequent property of refutability, as vital for both particular grammars and grammatical theories.

It is worth noting that the desideratum of predictive restrictiveness is closely allied to Popper's (1959) theory of scientific development. Popper claims that unless there is empirical evidence to distinguish among them, the best best of two competing hypotheses is the one which is compatible with the smallest number of distinct predictions. In other words, the more restrictive hypothesis is the better one[4].

An argument for **Univ-Fact** on the grounds of empirical restrictiveness might be formulated as follows.

(B) The Argument from Restrictiveness

(B-1) **Univ-Fact** cannot be proven empirically.

(B-2) **Univ-Fact** is not falsified by current evidence.

(B-3) **Univ-Fact** is more restrictive than **NoUniv**.

(B-4) More restrictive unfalsified hypotheses are to be preferred.

∴ It is better to assume **Univ-Fact** than **NoUniv**.

There is no problem accepting assumption (B-1); it is, after all, what we saw proven in the previous section. We shall for the purposes of this argument presume that (B-2) is also true. While (B-4) suffers some serious problems, we shall not tackle these here, but rather focus on assumption (B-3).

The counterargument to (B-3) is not direct. It relies on assuming the validity in general of arguments from restrictiveness. An alternative hypothesis is shown to be more restrictive than **Univ-Fact**, and so preferable to it. Furthermore, so long as this alternative is held, **Univ-Fact** lacks all restrictive power, and so the argument from restrictiveness can offer it no support.

3.1 Univ-Fact is restrictive

We begin by showing that **Univ-Fact** is restrictive. This is important, not so much for the result itself, but for what the argument shows about *how* **Univ-Fact** is restrictive. The reader should recall that we are only considering standard OT in which constraints are assumed to be subject to a total ranking: constraints cannot enter disjunctive relationships.

Suppose that some language uses two forms x and y for the same lexical input[5] in free variation[6]. As languages use single, fixed hierarchies, then this free variation must result from the equal harmony of these two candidates: they form a *tie*.

Now suppose that for some lexical input in a second language **Gen** produces a candidate set which contains both x and y. In this language, however, only candidate x surfaces as optimal. There is only one possible conclusion in this circumstance. The two languages must be employing different constraint sets.

[4] As an example, imagine that you have tossed a coin of unknown reliability 1000 times, and each time gained heads. One can imagine three hypotheses about the coin's behaviour: it always returns tails; it always returns heads, it returns anything. The first contradicts the data and is so eliminated. Of the remaining two, the second is more restrictive and so is preferred.

[5] For example, in my idiolect, /plant/ and /plænt/ occur in free variation as realisations of <**plant**>.

[6] Markus Walther (p.c.) suggests that this presupposition could not hold if the lexical entry were derived by lexical optimisation. If this is the case then the **Single** given below always holds, and **Univ-Fact** is not restrictive.

Because x and y are both optimal in the first language, they must incur precisely the same number of violations for each constraint in force in this language. If one of them is optimal in the second language, while the other is not, this means that they must incur a different number of violations for at least one constraint. It follows therefore, that the two languages cannot be using the same constraint set.

It is worth noting what knowledge we have assumed to be accessible in order to create this falsification of **Univ-Fact**; these are the same assumptions we made in section 2. We have assumed that we could identify two distinct surface forms which differ phonologically. This presupposes that phonological differences could be isolated from differences in phonetic implementation.

Secondly, we have presumed that it is possible to tell whether a second language uses these same candidates in a lexical competition. This is particularly difficult if, as in $\mathrm{OT}_{\mathrm{2\text{-}L}}$, candidates carry considerable non-surface structure with them. It may be that the second language has candidates with the same phonetic structure as the two optimal forms in the first language, but which carry different hidden structure, allowing the common constraint set to evaluate them differently.

These difficulties notwithstanding, there are circumstances in which the evidence we have assumed to be potentially available could falsify **Univ-Fact**. Thus it is, albeit in theory, an empirically restrictive hypothesis.

3.2 Univ-Fact is restrictive only with free variation

It so happens that **Univ-Fact** can only be falsified if there is a language showing free variation between two forms. We can show this by assuming the condition fails, and then proving that if any OT analysis of a set of languages is possible, then one respecting **Univ-Fact** is also.

Suppose we are examining a set of languages, and we analyse each of them using a different set of constraints. In each language, however, there is no free variation: from each candidate set, only a single optimal candidate is returned. For ease of reference, we will give this restriction a name, **Single**.

> **Single**: For every possible candidate set which **Gen** can output, the constraint hierarchy in each language selects only a single optimal candidate.

It should be noted that speakers of languages conforming to **Single** may realise the same word in a number of different ways, so long as the differences are ascribed to either phonetic implementation or separate lexical choice. **Single**, as used here, only deals with phonological constraint systems.

If **Single** holds throughout our analyses of each language, then we can construct another hierarchy for each language obeying both **Single** and **Univ**. This is done by appending to each constraint hierarchy all of the other constraints used in the analyses of the other languages. As these are lower ranked than the original constraints, the latter have priority in making their unique selection from the lexical candidate sets. The 'foreign' constraints are only able to select from within a singleton set of candidates. Thus for every input to **Gen**, the new, augmented hierarchies select the same optimal candidates as the original hierarchy. The action of the constraint hierarchies are precisely the same in the old and new analyses.

But notice that all of the new hierarchies use the same constraints. Consequently, they adhere to **Univ-Fact**.

So any OT analyses of any languages can be modified to conform to **Univ-Fact** without changing their empirical behaviour, so long as none of the original analyses allowed multiple winning candidates. In other words, any falsification of **Univ-Fact** must also falsify **Single**.

3.3 A more restrictive alternative

The problem for **Univ** is not that it lacks restrictive power, but that other hypotheses have more. In fact, **Single** is more restrictive, and by the argument from restrictiveness should be preferred.

We have already seen that if linguistic data from a number of languages allows **Single** then it also allows **Univ-Fact**. **Single** is therefore at least as restrictive as **Univ-Fact**. It is easy to imagine sets of languages analysable with the same set of constraints in different rankings, but which permit more than one optimal phonological form per lexical input. These would contradict **Single**. So while every contradiction to **Univ** is a contradiction to **Single**, the reverse is not the case. Thus **Single** is more restrictive than **Univ**.

In the argument from restrictiveness, assumption (B-4) bids us prefer unfalsified restrictive hypotheses. Until **Single** is disproven, it should be preferred to **Univ**.

3.4 Both hypotheses together

Single and **Univ** are not incompatible. If they are both restrictive, is there not a case that both be accepted? The answer is that while they can be entertained simultaneously, the argument from restrictiveness provides no reason why they should be.

We saw above that **Single** is disproven by all counter-examples to **Univ-Fact**, and more. Counter-examples to the conjunction of the two constraints will be the union of two counter-example sets. This will be identical to the counter-example set for **Single**. So **Single** and **Univ-Fact** combined are no more restrictive than **Single** on its own.

In summary, then, if empirical restrictiveness is to be a desideratum, then it is not one which **Univ-Fact** maximises. In fact, a more restrictive hypothesis, **Single**, robs **Univ-Fact** of any restrictive power. Until **Single** can be convincingly falsified, the argument from restrictiveness offers no support to **Univ-Fact**.

The next section examines whether simplicity can offer any support to **Univ-Fact**.

4 Simplicity

The third argument for **Univ** departs from empirical considerations and resorts to that most powerful of non-empirical arguments: simplicity.

(C) The Argument from Simplicity

(C-1) **Univ** is simpler than **NoUniv**.

(C-2) Simpler hypotheses should be preferred.

$\therefore$ **Univ** should be preferred to **NoUniv**.

This argument both succeeds and fails. It succeeds when **Univ** is interpreted as a convention, i.e. when **Univ** is **Univ-Conv**. But when **Univ** denotes a fact about psychological reality, assumption (C-1) fails, and the argument consequently lends no support for **Univ-Fact**.

The one successful simplicity criterion requires models to be evaluated in their complete description. The simplicity of a hypothesis cannot be evaluated independently of how it affects the representation of the data it accounts for, the probability of phenomena it is to explain, or the ad-hocness of the theoretical infrastructure it entails. The feature-counting measures of simplicity of phonological analyses were examples of this kind of simplicity measure precisely when they balanced the complexity of rules systems against the elimination of redundancy from the lexicon.

In fact, this complete view of simplicity defines the machine-learning method known as minimum message-length (Wallace & Boulton 1968, Wallace & Freeman 1987) or minimum description length (Rissanen 1978, 1982, 1987) which has close links to Bayesian probability and algorithmic complexity (Li & Vitanyi 1989, 1993). The same method has been used to computationally select between different phonological analyses (Ellison 1992). It behoves us then to evaluate the simplicity of **Univ** together with the theoretical edifice it entails.

4.1 The simplicity of Univ-Conv

Let us first consider the simplicity argument for **Univ-Conv**. In this case, we need only consider the analysis as a description. Descriptions have neither a causal effect, nor do they need to be explained as the result of particular causes. Rather, they are formally self-contained. Thus it is only the components of the analysis itself which needed to be measured in a simplicity argument.

An OT description of one language needs five components: lexical inputs, **Gen**, a list of constraints, a ranking of the constraints and **Eval**[7]. Being willing to assume the universality of **Gen** and **Eval**, for the purposes of this argument, we do not need to express each of these components anew for every different language. A description of the, say, six thousand languages of the world consequently needs less than thirty thousand components. Grossly measured, for six thousand languages, we need 18002 components: **Gen**, **Eval**, 6000 sets of lexical inputs, 6000 constraint sets and 6000 constraint rankings. This is the component count if **NoUniv** is adopted and languages vary in their constraint sets. Does **Univ-Conv** make things simpler?

If the same constraint set is used in each description, then we need only 12003 components: **Gen**, **Eval**, the universal constraint set, 6000 sets of lexical inputs and 6000 constraint rankings. So the comparison is between 18002 components, or 12003. This is not a definitive proof that **Univ** is better. It may be a comparison of 18002 simple, transparent components with 12003 components of fiendish complexity. Ceteribus paribus, however, it is reasonable to assume that analyses with fewer components are simpler.

[7] We assume that the expression of the constraint set, e.g. as a bitmap, does not imply any rank ordering.

So **Univ-Conv** is well supported by the argument from simplicity. Its realist counterpart **Univ-Fact**, however, falls foul of the need for explanatory causes.

4.2 The simplicity of Univ-Fact

The objective components needed for any OT model of the cognitive language processes include those needed for language description: lexical inputs, **Gen**, constraints, constraint rankings and **Eval**.

A principle emphasised by Isaac Newton (1953) and more recently by Reichenbach (1956), Salmon (1975, 1978, 1984) and Sober (1988) requires that to the same natural effects common causes should be assigned. In its more modern formulation, correlations should be explained by means of a common cause. Salmon offers the example of word-for-word identical assignments being submitted by two students. It is possible they were created independently; it is more plausible that at least one of the students is guilty of plagiarism.

This principle applies to universal tendencies in language structure: they need to be explained by a common cause, which can justify the assumption of a universal **Gen**, a universal **Eval** and a universal constraint set. One possible such cause will be discussed in more detail in section 5. For the current argument, the important implication of this principle is that if we assume that each language speaker's mind embodies the same **Gen**, **Eval** and constraint set, then we must present common causes for these to account for their uniformity.

As we are here only interested in the contrast between **Univ** and **NoUniv**, we will presume that explanations are found for universal **Gen** and **Eval**. So we need only seek a common cause for the uniformity of the constraint set across humanity.

This need for a common cause arises with the constraint ranking as well. Speakers of the same dialect use the same ranking. Such a correlation needs to be explained. The explanation given in OT is that the linguistic environment provides evidence, and a learning procedure reranks constraints until the adult, correct ranking is achieved. So the OT model of the language user involves linguistic input and a reranking procedure.

Univ-Fact applies not only to adults but to children as well. It follows that children at all ages must in fact share the same constraint set as adults. Consequently, the common constraint set cannot be acquired. It must be innately, presumably genetically, specified.

Many other domains, apart from language, can be modelled by an OT-like system of ranked constraints. For example, driving a car may be reduced to a number of constraints, some of which take priority over others. These constraints interact to select optimal actions for the driver. Highly ranked will be *don't hit anyone*, more lowly ranked will be *go as fast as you can*.

While some of these constraints might be universal, and be shared with many other skills, such as *don't hit anyone*, others, like *depress the clutch before changing gear*, will not. Since we learn to drive vehicles nonetheless, the human mind requires a mechanism for learning such constraints. Now let us return to language.

If the phonological constraint set is innate, then it cannot be the result of learning, and so a cause beyond that used to acquire constraints for other skills, e.g. driving, must form part of our model of cognitive development and function, adding to its complexity.

We can evaluate **Univ-Fact** in terms of the complexity of the model it requires for the language user. **Univ-Fact** requires that the speaker begin with the following objects: (i) **Gen**, (ii) **Eval**, (iii) linguistic input for learning the lexical inputs and the constraint ranking, (iv) a mechanism for ranking constraints, (v) a mechanism for learning constraints in other domains and (vi) a genetic stipulation of the common constraint set.

Without **Univ-Fact**, we can have a simpler model because all constraints can be created by the same mechanism, a learning device. The components needed by the model are: (i) **Gen**, (ii) **Eval**, (iii) linguistic input for learning the lexical inputs and the constraint ranking, (iv) a mechanism for ranking constraints and (v) a mechanism for learning constraints in all domains. This model is simpler than that needed for **Univ-Fact**.

Of course, even without adopting **Univ-Fact**, we could propose that constraints were determined genetically. Different gene combinations would be needed to account for different constraint sets. At first glance, this might appear as complex as having a genetic specification for a universal constraint set. The principle of similar effects having similar causes means, however, that the uniformity of constraints under **Univ-Fact** must be the result of uniform genetic specifications. But why should the genetically specified constraint set not vary genetically, as does body-shape, eye-colour or fingerprints? A further causal mechanism is needed to explain the uniformity of the genetic specification for the constraint set.

So the genetic specification of the universal constraint set does no more than move the uniformity under contention from the speakers' minds to their genes. The need for a common cause explanation to account for **Univ-Fact** remains.

When placed in its complete setting, **Univ-Fact** makes for a more complex model of language than does **NoUniv**. Uniformity in the real world is a simplifying assumption only so long as it can be attributed naturally to an otherwise motivated common cause. If the uniformity comes at the expense of added ontological assumptions, such as a relatively uniform genetic specification of the constraint set, then it makes for a more complex, not a simpler, assumption.

In summary then, we have two very different evaluations of **Univ**. The factual version **Univ-Fact** is not well supported by a simplicity argument, in fact, simplicity favours **NoUniv**. But in its application to phonological descriptions, **Univ-Conv** seems a natural step towards constructing simpler and more concise simultaneous analyses of many languages, without the burden of ontological claims.

5 Markedness and Acquisition

The next two arguments for the universality of constraints are treated together for two reasons. Firstly, they have much in common. Both rely on **Univ-Fact** as necessary for the explanation of certain phenomena. Secondly, a single response counters both arguments.

The first of these two arguments for **Univ-Fact** is the argument from cross-linguistic markedness. It is based on the observation that certain phonological structures seem to be preferred in all languages.

(D) The Argument from Cross-Linguistic Markedness

(D-1) Languages regard the same structures as unmarked.

(D-2) **Univ-Fact** offers an explanation of this.

(D-3) There is no other explanation of this.

∴ **Univ-Fact**.

One example of markedness, in the sense of (D-1), concerns voiceless stops. Languages generally have either both voiced and voiceless stops, or only voiceless stops. No language uses only voiceless stops. Voiceless stops are therefore said to be unmarked.

Similarly, all languages seem to use /CV/ syllables even though they may have more elaborate syllable types as well. While there are languages which allow only this kind of syllable and no other, there are no clear cases of languages prohibiting it.

The second of the two arguments for **Univ-Fact** has the same form as (D), but addresses the order in which particular linguistic constructions are acquired. **Univ-Fact** is needed to account for the correlation between the order of acquisition of phonological structures and increasing cross-linguistic markedness.

(E) The Argument from Acquisition Ordering

(E-1) Children learn to produce more marked phonological structures later.

(E-2) **Univ-Fact** offers an explanation of this.

(E-3) There is no other explanation of this.

∴ **Univ-Fact**.

The basis of this second argument, namely assumption (E-1), can be termed the Jakobsonian Generalisation after the linguist who first stated it (Jakobson 1968). It has been pursued in linguistic theory by Stampe (1979), and recently, Smolensky (1996).

5.1 Explanations with Univ-Fact

Optimality theory with **Univ-Fact** offers an account for both (D-1) and (E-1). The account of (D-1) is quite simple: markedness is equated with constraint violation. All languages share the same constraints. The least marked forms in any language will be the ones which incur no violation to any constraint. Forms which violate no constraints will also be optimal in all other rankings of the same constraints, and thus, under **Univ-Fact**, be universally unmarked.

Smolensky (1996) offers an $OT_{2\text{-}L}$ account of (E-1). His account relies on the demotion algorithm for learning constraints developed with Tesar (1995, Tesar & Smolensky 1993, 1996). Boersma (this volume) offers an alternative algorithm for the same task. These algorithms require the learner to know *a priori*, or deduce, the lexical input as well as the complete surface representations of words it hears. With the lexical input, **Gen** can be used to construct the lexical constraint set.

If the correct candidate, that is the form actually appearing in the language, is ruled non-optimal in the current ranking, all constraints which prefer other candidates to it are demoted below the highest-ranked constraint which will eliminate these competitors. This algorithm can be proved to arrive at a ranking which selects as optimal the right surface forms, provided such a ranking exists. Smolensky proposes that children learn language by demoting constraints in this way.

In order to explain (E-1) with constraint demotion, two problems must be solved. First, how does the child determine the lexical input for new words it hears? Second, what relates phonological markedness to acquisition order?

In answer to the first question, Smolensky stipulates that the learner treat the perceived form as the lexical input. **Gen** acts on this form, pairing it with all possible surface forms. If the only constraints which were to apply were faithfulness constraints, then the child would reproduce the input exactly.

The second question is also answered by stipulation. The learner begins with a constraint hierarchy which ranks all phonological structure constraints above all faithfulness constraints. As the child receives input which conflicts with phonological markedness preferences, the structure constraints are demoted, and the undemoted faithfulness constraints effectively percolate higher in the constraint ranking. If, however, the child never receives input which contradicts a particular well-formedness constraint, such as a child raised in a Hawaiian-speaking environment will not hear codas, then the constraint is never demoted and so remains to outrank faithfulness constraints. In maturity, the hierarchy will result in borrowed lexical items being realised in conformance with the undemoted phonological structure constraints.

That **Univ-Fact** contributes to an explanation of these two phenomena offers little support for **Univ-Fact** if there are simpler alternatives. The next section provides one such alternative.

5.2 Acquisition without Univ

The starting point for a universal-less account of markedness and acquisition order is the common human development in physiology and coordination. Very young infants face two problems in the production of the words they hear[8]: the shape of the mouth makes some segments impossible to produce, and their lack of general coordinative skills also preclude the reliable production of some segments and/or sequences (Kent 1992a, b, Kent & Miolo 1995:307-9).

During the process of development children outgrow their physiological limitations and, more selectively, the limitations on their ability to coordinate sounds. With the addition of two further assumptions, this provides the basis for the universal-less explanation. The first assumption we make is that children learn constraints which internalise the structures of those words which they say repeatedly. This assumption is shared by the psycholinguistic model of Menn (Kiparsky & Menn 1977, Menn 1983), and later Matthei (1989), in which developing children store their own utterances as well as those perceived in their linguistic input.

The second assumption is that children tend to use words which they know they can articulate successfully. While they continually try new ones, they do

[8] I leave aside the question of perceptual development here, although it must certainly form a component of any complete theory.

not continue to repeat words by means of which they have failed to communicate. Three different kinds of evidence can be adduced for this assumption. Firstly, in their earliest meaningful utterances children tend to reuse syllables used in babbling (Vihman 1992). Secondly, at a later age, children are found to actively avoid words and phonemes which they cannot produce accurately (Ferguson & Farwell 1975, Schwartz & Leonard 1982). Further, children spontaneously self-correct and repair following a failure to communicate (Clark 1978).

From these two assumptions, we conclude that a child at any stage of development will be revising the phonological constraint system to account for the intersection of the language it is exposed to, and the capabilities it has within the physiological and coordinative limitations of its developmental stage. While these limitations may be quite similar between children, there is no evidence or necessity to suppose that the constraints which the child uses to internalise these limitations do *not* vary significantly from child to child. It is the function of the constraint system as whole, not of its components, which matters.

It is this internalisation of the child's own articulatory limitations which accounts for the broad similarities in cross-linguistic markedness judgements. But to complete this account, we need the assistance of one final assumption: that children retain permanently the linguistic knowledge gleaned at previous stages of development. Later development only adds to this knowledge, supplementing the constraints which modelled the limitations on structure and articulation present at earlier stages.

On the basis of this assumption, the child always retains the constraints which modelled its articulatory skill at earlier stages of development. These constraints may be outranked by later-developed constraints, constraints which are perhaps similar to the faithfulness constraints in $OT_{2\text{-}L}$. In any case, the most optimal word forms will be those which conform to all of the constraints, including those learnt during the period of limited articulatory prowess. Thus the least-marked utterances will be those that were possible at the earliest stages of physiological and coordinative development.

Where constraints describing outgrown limitations are outranked by new constraints developed in response to linguistic input, the child learns to articulate more marked structures. Where they are not, the limitations of child physiology and coordination are reflected in the corresponding aspects of the adult pronunciation. The adult may remain incapable of articulatory feats not because their tongue lacks agility, but because the model of lingual articulation they have internalised has not been revised since a less agile stage of development.

For example, consider a child attempting to produce the word /snɔʊ/, *'snow'*. When it first attempts speech, the child will find itself unable to reproduce the initial consonant cluster of the word, producing perhaps [nɔ] or [soʊ] instead. It will internalise this limitation in its cognitive model of phonology. At some point in development, the child will attain both the physical and coordinative ability to say the cluster /sn/. But more is required, namely continued pronunciation of such clusters by the child. The impetus for the pronunciation of certain clusters is present in the language: if the child pronounces a word with an onset cluster, and succeeds in better mimicking adult pronunciation, then this behaviour is reinforced, and the child is likely to use that pronunciation again. With continued use, the child reviews its cognitive model of the articulation system to account in a comprehensive way for the forms it now finds itself saying.

In OT terms, this review of the articulation model need be no more than the addition of new constraints. There is no need for reranking of the constraints describing earlier stages. The child learning to pronounce complex onsets, may dominate its previous hierarchy with a constraint which forces the realisation of the perceived sequence /sn/ as a complex onset. Even if all earlier models of the articulation system had precluded consonant clusters, the dominance of this constraint will allow clusters to surface where lexical forms contain them.

So we now have an account of both phenomena, (D-1) and (E-1), which does not depend on **Univ-Fact**. This account also does not need the mechanism of constraint demotion. Rather, all change is effected by the construction of new constraints which may override the constraints reflecting earlier, developmental limitations. In this account we have used as causes only linguistic input, a learning device, and the child's own physiological and coordinative development. In any OT account of language development using **Univ-Fact**, these facts remain, but must be supplemented by an added mechanism to force constraint uniformity. As pointed out in the discussion of simplicity in section 4, this kind of additional requirement renders the explanation from **Univ-Fact** more complex.

Taking the universal-less view to its limit, we might assume that once an initial set of well-formedness constraints are learnt, further constraints specify morpheme classes, or in the extreme case, individual morphemes in the lexicon. The adult constraint system, after such a learning process, would look like that proposed by Russell (1995). Russell contended that rather than being the inputs to **Gen**, lexical specifications are constraints which limit a single universal candidate set. These morphemic constraints can be ranked among the phonological constraints, and it is the interaction of the two that produces morphophonological complexity.

In this model, complex, later-learnt articulation patterns such as morpheme-specific phonological effects are reified in constraints of higher rank. This is appropriate as these phenomena are supplementary cases of the *'except when'* behaviour which motivated much of Prince & Smolensky's (1993) argument for OT[9].

In summary, then, this section has presented an alternative account of both cross-linguistic markedness effects and the correlation of markedness with acquisition order. This account, while making use of the totally ordered constraint hierarchy of OT, does not assume a universal constraint hierarchy. Instead, constraints are learned to reflect the growing articulatory capabilities of the child and the demands of the language being learned. These constraints are not given but made.

As this alternative account requires only the staged development of the child from a common, very limited, state, and does not require wholesale stipulation of linguistic, non-physiological universals, it is simpler than explanations built from **Univ-Fact**, and so forms a preferable explanation. Consequently, the arguments for **Univ-Fact** from markedness and from acquisition have little weight.

[9] Thanks to Markus Walther (p.c.) for pointing this out.

6 Learnability

The penultimate argument for **Univ** is the argument from learnability. This argument is one shared, at least in part, with other theories of linguistics which rely on the notion of Universal Grammar.

(F) The Argument from Nativism

(F-1) Adults speakers use constraint hierarchies to define their language.

(F-2) These hierarchies must have come from somewhere.

(F-3) Learning constraints and hierarchies from positive data alone is in general too difficult for children too accomplish.

(F-4) Children only have access to positive data.

$\therefore$ The constraints are not acquired.

(F-5) What is not acquired is innate.

$\therefore$ The constraints must be innate.

(F-6) Humans are all equally capable of learning all languages.

(F-7) If humans had different innate constraints, they could not learn all languages equally well.

$\therefore$ All speakers have the same innate constraints.

(F-8) All languages use all constraints.

$\therefore$ **Univ-Fact** All languages use the same constraint set.

The view that universal grammar is the result of genetically specified mental structures specific to language is widely held in the linguistic community (for a recent popular exposition, see Pinker (1994)). This view has, however, recently come under attack from connectionism (Elman et al. 1996, Quartz & Sejnowski 1996) and statistical learning (e.g. Finch & Chater 1992).

The argument above is an example of how the universality of the constraint set might be linked with the innateness. If we need the genetic specification of universal grammar to make languages learnable, then it is reasonable to assume that the genes specify the constraint set. If this is so, then all languages will use the same constraint set in the same way that (almost) all people are born with four fingers and a thumb on each hand.

There are, however, three points where this argument can be challenged: assumptions (F-4), (F-7), and (F-8). The remainder of this section discusses problems with each of these assumptions in turn.

6.1 Poverty of the stimulus (F-4)

The first of these three assumptions states that children only have access to positive information, that is, information about what is *possible* in the language they hear. In contrast, they receive no information which indicates that a certain form or construction is *impossible* — they receive no negative evidence.

This premiss is based on the evidence that children do not seem to receive care-taker instruction that ungrammatical utterances they make are improper. On the occasions that they do receive this kind of input, they seem to ignore it (Brown & Hanlon 1970 , Pinker 1989). This view of child input has not, however, gone unchallenged (Sokolov & Snow 1994).

One form of functional negative evidence which is available to the language acquirer is failing to communicate. If a child asks for an icecream, and is met with blank stares, this offers significant evidence that the construction failed. Reasons for the failure may be pragmatic, lexical or grammatical. If there is other evidence to show that the pragmatics and lexicon are satisfactory, then grammatical infelicity is the likely cause of the failure to communicate. Negative evidence has been gleaned about the grammar.

It would intuitively seem to be the case that the lack of a particular construction in the ambient linguistic input could also act as negative evidence. However, it is argued in the literature that this is not the case (Valian 1990). The reason for this is not so much the nature of language, as the nature of current linguistic models.

Most current models of language do not regard information about the frequencies of items or structures as part of the systematic specification of the language. Consequently, the incorporation of a particular construction in a language gives no indication, and in fact, *can* give no indication of how frequent that construction is. Consequently, its lack in available data may merely result from the construction having low frequency, not from its systematic prohibition.

Note that this ignorance of distributional evidence is a vital assumption in Gold's (1967) proof of the necessity of negative evidence for learning one of a suitably large class of languages.

In contrast, a language model of language which regards frequency information as part of the specification of a language can be subject to negative evidence, albeit not *absolutely* conclusive evidence, if a form which should occur in the language with relative frequency f does not occur. As the number n of constructions in which the form could occur but does not increases, the probability $(1-f)^n$ of not seeing an example of the form tends to zero. Bayes' theorem then implies that the likelihood of a grammar which predicts this frequency f for the form must also tend towards zero.

In other words, the absence of expected constructions can act as negative evidence to the right sort of language model. Conversely, the lack of negative instruction only implies a lack of negative evidence if the language model is too impoverished to make substantive claims about frequency. There seems little evidence that children's model is so impoverished.

An alternative form of negative evidence occurs if the relations between language structures are topographic (see Ellison (1997) for a discussion of learning with topographic mapping, and its potential application to language). If a language has a topographic mapping from meaning to phonological form, then similar meanings, at, e.g. the sentence level, relate to similar phonological forms.

This offers implicit negative evidence in the following way. If meaning A is similar to meaning B, and the output form of A is a, then the output form of B is (probably) not dissimilar from a.

In Optimality Theoretic terms, this amounts to the restriction that similar lexical inputs should result in similar optimal candidates. For example, if the Old Irish lexical input /berami/ results in an optimal output form [bermai], then we would expect the input /gerami/ not to result in [graim] in preference to [germai]. This is because [germai] is more similar to [bermai] than is [graim].

If learners distinguish successful interactions from unsuccessful, if grammars specify distribution, or if grammars make use of topographic mappings, then negative evidence is available to language learners even without negative instruction. If the learner is using language to achieve a goal, then failure to achieve this goal may indicate an improper construction. If the learner has distributional expectations, or expectations based on similarity, failure of these may also indicate a need for grammatical revision.

Of course, the best response to this assumption of poverty of the stimulus would be to build a system which was capable of learning language structure without negative evidence, perhaps using distribution or topographicality assumptions to achieve this end. Unfortunately, approaching this problem is beyond the scope of this paper, and a final solution is still lacking.

In summary, then, the claim that children receive no negative evidence is definitely arguable, relying on isolating the task of learning about grammar from considerations of motivation, distribution and topographicality.

6.2 Innate constraints must be uniform (F-7)

Another assumption which can be challenged states that languages could only be learned with equal facility if all learners had the same innate constraint set. Let us leave for section 6.3 the possibility that learners need not use all of their innate constraints. Even without this option, it may be the case that the space of human languages is accessible using a number of different constraint sets.

It may be the case, for example, that variation is restricted to constraints which are applicable only in a very restricted range of situations. Variation in these constraints would offer no more handicaps to speakers than the variation which occurs in the inclusion of low-frequency words in our individual vocabularies.

For example, constraints controlling the fine prosodic interaction of words cross-clausally may be present in some people and absent in others. This will allow the former group a finer poetic ear than the latter. If other parts of linguistic behaviour have a genetic basis, there is no particular reason to think that this kind of individual difference cannot have a basis in genetic variation in the same way that eye colour, skin tone or height have.

Of course, these low-frequency constraints may be so outranked, that any effect they might have on the language of the individual is overshadowed by other influences. The individual differences offered by these constraints might then be curtailed by an appropriate conspiracy of shared constraints.

The important point to take from this section is that while uniformity of constraints may be important for the most frequently active constraints, it is by no means certain for the less-frequently applied constraints. For these cases, a lack of uniformity in the innate constraint set results in little communicative cost.

6.3 All languages must use all innate constraints (F-8)

Another assumption open to challenge is (F-8). This assumption claims that all the innate constraints must be used in all languages. Without it, even a shared human gene-complex specifying a uniform constraint set will allow different individuals and/or different languages to make use of distinct subsets of the common constraint set, and consequently, the constraint sets used in language hierarchies will not be universal. **Univ-Fact** will fail.

The argument for innateness presumes an inborn candidate set in order to make the task of language learning tractable. Language learning is the task of identifying the correct grammar in a large space of possible grammars. The argument for innateness claims we need a language-specific genetic endowment to reduce the space of possible grammars to a size at which it can be feasibly searched.

In OT, the space of grammars is the space of constraint hierarchies. The innate specification of **Con** reduces complexity by limiting possible hierarchies to those using a single common constraint set. This, it is argued, makes learning tractable, as it involves only the task of constraint reranking.

But the innate specification of a common constraint set need not force all languages to use the same constraints. Different languages might make use of different subsets of the constraint set. This could be achieved in two ways, with language development consisting of removing constraints from an all-encompassing initial state, or of adding constraints to a constraint-poor initial state.

In the former case, a child is born with a hierarchy which includes all of the innately possible constraints. As the child gathers linguistic evidence, certain constraints are eliminated, when they are seen not to participate in the hierarchy.

The most direct evidence for constraint deletion is free variation. Suppose that for a particular lexical input, two different surface forms are equally acceptable in the target language, and what is more, these two candidates differ only in the evaluation of one constraint. As no hierarchy, stratified hierarchy, or even hierarchy with disjunction, can account for this kind of free variation, it follows that the constraint which distinguishes the two forms cannot be part of the hierarchy defining this language.

As a concrete example, let us imagine that syllable structure is defined by the constraints **Ons**, **NoComplex**, **NoCoda**, **Parse** and **Fill**. Suppose free variation occurs in which an high vowel followed by a low vowel can be parsed as the head of its own syllable, or as an onset to the following vowel. These two cases might be /<u>i</u>.<u>a</u>/ and /i<u>a</u>/. In the first case, **Ons** is violated twice, but no other constraint is violated: there are no complex onsets or codas, and there is no deletion or epenthesis. The same is true of the second syllabification except that **Ons** incurs no violations; the only syllable has an onset.

If these two syllabifications occur in free variation, then **Ons** cannot form part of the hierarchy, or the variation would not be free. No reordering, or even disjunction, of these constraints can account for this variation. Thus **Ons** cannot form part of the constraint hierarchy for this language.

It does not matter that in a different hierarchy, this free variation might be analysed by constraint disjunction. Nor is it important that this kind of free variation might not actually occur. The point is that a single example of free variation is enough to eliminate an unwanted constraint from a hierarchy, without any great computational expense. Subset hierarchies are, therefore, learnable.

We might also choose to remove constraints which never serve to eliminate non-optimal candidates. For example, **Ons** might be removed in a language with a highly ranked **Parse** dominating **NoComplex** which in turn dominates **NoCoda**, simply because these three constraints combine to force any intervocalic consonant to be parsed as an onset. If the language does not permit syllabic consonants, then the only alternatives to parsing a consonant as an onset are deletion or parsing it as a coda. These two alternatives are precluded by the highly ranked **Parse** and **NoCoda** constraints. Thus eliminating the **Ons** constraint would have no effect on the syllabification of lexical material. The constraint could then be deleted.

That **Ons** *can* be dispensed with does not mean that it necessarily is removed from the hierarchy. What it does mean, however, is that if the learning algorithm does dispense with the constraint, this will not affect the language of the learner. **Univ-Fact** is not required to make language learning feasible.

In an alternative model, constraints are inserted into a hierarchy rather than removed from it. Suppose we begin with a hierarchy containing only structural constraints, and no faithfulness constraints. Whenever the constraints in the hierarchy are violated, and reranking will not alleviate the problem, a faithfulness constraint could be introduced from the innate constraint set which requires just enough faithfulness in the input-output mapping to make the correct candidate optimal.

The new constraint could subsequently be ranked within the hierarchy using an algorithm like that of Boersma (this volume) or of Tesar & Smolensky (1996).

There is an interesting implication of this model which distinguishes it from accounts, like Smolensky's (1996), in which faithfulness constraints are present but ranked low in the initial stage of acquisition. In both models, early articulations will be largely the effect of structural constraints. The difference between the models is that having faithfulness constraints within the infants grammar from the beginning means that upto the limits of their linguistic prowess, children will always attempt to produce real, well-formed words.

In contrast, if children begin with no faithfulness constraints in the initial stages, then we might expect an initial stage of unmarked and also meaningless articulation: babbling. This is what does occur in babies. Babbling is devoid of word content. This supports the proposal that faithfulness constraints are absent, rather than dominated.

In summary, then, babbling offers some evidence that constraints are inserted into the constraint set as the child language learner develops.

At issue in this section is not so much the contingent fact of whether or not all of the putative innate constraints are used in all languages, but the necessity of this being the case. Given that procedures to insert or delete constraints are not computationally taxing, learnability offers no basis for assuming this consistency.

6.4 Summary

We have looked at three assumptions used in the argument for universality from learnability via innateness. The poverty of the stimulus premiss (F-4) can be countered by incorporating motivational, distributional or topographic information into the language model.

The assumption that all innate constraint sets must be the same (F-7) can also be challenged: even if the difficulty of learning forces us to presume that constraints are innate, there is no necessity that we all have the same innate set. All that is required for the relative homogeneity of our linguistic capabilities is that variation predominately occur in constraints with a low frequency of application, much as variation in native speaker vocabularies occurs primarily in the low frequency items.

Nor is the third assumption, that all users must incorporate all innate constraints into their hierarchies (F-8), a self-evident truth. Identifying situations in which constraints can or must be deleted from hierarchies is computationally tractable, as is identifying when they should be inserted. Furthermore, an insertion model makes the correct prediction that the initial stage of language should be arbitrary babbling, not merely the reduction of meaningful forms to unmarked articulations.

The conclusion, then, is that the argument for innate language knowledge from the poverty of the stimulus premiss, does not of itself offer significant support to the claim that the constraint hierarchies of the world's languages use the same constraint set.

7 Convention

The final argument for **Univ** is the most powerful.

(G) The Argument from Ease of Communication

(G-1) Using a uniform constraint set makes it easier for phonologists to communicate analyses of languages to each other than if they used different constraints for each language.

(G-2) The easier it is for phonologists to communicate language descriptions, the better.

$\therefore$ Phonologists should use a standard constraint set in analysing languages.

It is very difficult to contest either assumption in this argument. If language descriptions are sought using a common constraint set, and individual descriptions differ only in the rankings of the constraints, then a linguist can quickly grasp the distinctive content of new language descriptions. This is particularly so if they are already familiar with a number of different rankings of the same constraints.

On the other hand, if the burden of understanding the grammar of another language includes mastering the implications of a new set of constraints *as well* *as* grasping the implications of the constraint ranking, then the task will be much more difficult[10].

This argument, then, seems sound. A standard set of constraints makes a useful tool for furthering communication between linguists.

8 Conclusion

In the introduction to this paper, I set out to show that **Univ** made a better tool than fact. Sections 2 to 6 discussed arguments which sought to establish **Univ** as a fact, through empirical means, restrictiveness, simplicity or as a necessary explanation for features of markedness, acquisition order, or learnability. Each time, **Univ-Fact** as fact proved to be an escapable conclusion. In contrast to these arguments, the case for **Univ** as a conventional usage to make the sharing of linguistic descriptions straightforward is robust. The conclusion is that **Univ** is not a fact, but a promising convention.

The international phonetic alphabet (IPA) makes an excellent parallel. More than any theory or linguistic fact, this convention has allowed each conforming linguistic description and analysis access to a wider audience. A conventional set of constraints for language description, independent of any theory-particular claim of universal grammar, cross-linguistic markedness or mental reality, would serve similarly to make language descriptions more accessible, more understandable, and more readily matched against theoretical speculations.

If **Univ** is taken as a convention, then our attitude towards it can be more flexible. For example, requiring that all constraints are present in all languages is as useful as requiring that every language employ all the sounds tabulated in the IPA. Rather, the common constraint set becomes a resource for language description from which the linguist can draw according to their needs.

Secondly, if the phonologist needs a constraint not found in the universal constraint set, there is no need to create another grand concept to be found in all languages. Instead, a diacritical rider on an existing constraint, like diacritics on phonetic symbols, will make easier the communication of the phenomena. This is particularly so if the standard constraint set comes replete with standard riders. These might include *except at the beginning of a word*, or *in open syllables*[11].

OT offers a remarkable opportunity. For the first time, the linguistic community can define a common language for phonological generalisations akin to the phonetic alphabet. For the first time, the means of combining phonological generalisations is sufficiently flexible to allow the same ones to analyse many different languages. But we should shy away from making a category error, confusing description with content, alphabet with inventory. **Univ** makes a rich device, but a poor fact.

[10] Markus Walther has noted (p.c.) that if declarative constraints are used, the problem is also simplified: no rank ordering needs to be considered, and the independent action of declarative constraints makes it possible to understand each in isolation.

[11] This does not have the same effect as ranking another constraint higher or lower. Applying such riders to a constraint would have the effect of making the whole hierarchy more permissive, rather than less.

References

Antworth, E. (1990). *PC-KIMMO: A Two-Level Processor for Morphological Analysis.* Dallas TX: Summer Institute of Linguistics.

Archangeli, D. (1997). Optimality Theory: An Introduction to Linguistics in the 1990s. In D. Archangeli & D. T. Langendoen (Eds.), *Optimality Theory: An Overview* (pp. 1–32). Blackwell.

Archangeli, D. & Langendoen, D. T. (Eds.). (1997). *Optimality Theory: An Overview.* Blackwell.

Archangeli, D. & Pulleyblank, D. (1989). Yoruba vowel harmony. *Linguistic Inquiry, 20,* 173–217.

Archangeli, D. B. & Pulleyblank, D. G. (1994). *Grounded phonology.* Number 25 in Current studies in linguistics series. Cambridge, MA: MIT Press.

Benua, L. (1995). Identity effects in morphological truncation. In J. Beckman, L. W. Dickey, & S. Urbanczyk (Eds.), *Papers in Optimality Theory,* number 18 in University of Massachusetts Occasional Papers in Linguistics (pp. 77–136). Amherst, MA: Graduate Linguistic Student Association. Rutgers Optimality Archive, ROA-74, http://ruccs.rutgers.edu./roa.html.

Bird, S. (1990). *Constraint-Based Phonology.* PhD thesis, University of Edinburgh. Revised as Bird 1995.

Bird, S. (1995). *Computational Phonology: A Constraint-Based Approach.* Studies in Natural Language Processing. Cambridge University Press.

Bird, S. & Ellison, T. M. (1994). One level phonology: autosegmental representations and rules as finite automata. *Computational Linguistics, 20,* 55–90.

Brown, R. & Hanlon, C. (1970). Derivational complexity and order of acquisition in child speech. In J. R. Hayes (Ed.), *Cognition and the Development of Language* (pp. 11–53). New York: Wiley.

Chomsky, N. (1978). A Theory of Core Grammar. *Glot, 1,* 7–26.

Chomsky, N. & Halle, M. (1968). *The Sound Pattern of English.* New York: Harper and Row.

Clark, E. V. (1978). Awareness of Language: Some Evidence from What Children Say and Do. In A. Sinclair, R. Jarvella, & W. Levelt (Eds.), *The Child's Conception of Language* (pp. 17–43). New York: Springer.

Dickey, M. (1995). Inversion in child English and acquisition in Optimality Theory. In J. Beckman, L. W. Dickey, & S. Urbanczyk (Eds.), *Papers in Optimality Theory,* number 18 in University of Massachusetts Occasional Papers in Linguistics (pp. 575–588). Amherst, MA: Graduate Linguistic Student Association.

Ellison, T. M. (1992). *The Machine Learning of Phonological Structure.* PhD thesis, University of Western Australia, Perth.

Ellison, T. M. (1997). Induction and Inherent Similarity. In Ramscar, M. & Hahn, U. (Eds.), *Proceedings of Simcat'97,* (pp. 83–90).

Elman, J. L., Bates, E. A., Johnson, M. H., Karmiloff-Smith, A., Parisi, D., & Plunkett, K. (1996). *Rethinking Innateness.* Cambridge MA: MIT Press.

Ferguson, C. A. & Farwell, C. B. (1975). Words and Sounds in Early Language Acquisition. *Language*, *51*, 419–439.

Finch, S. P. & Chater, N. (1992). Bootstrapping Syntactic Categories Using Statistical Methods. In W. Daelemans & D. Powers (Eds.), *Background and Experiments in Machine Learning of Natural Language* (pp. 229–236). Tilburg, NL: ITK.

Gold, E. M. (1967). Language Identification in the Limit. *Information and Control*, *16*, 447–474.

Golston, C. & Wiese, R. (1996). Zero morphology and constraint interaction: Subtraction and epenthesis in German dialects. *Yearbook of Morphology*, *1996*, 115–142. Rutgers Optimality Archive, ROA-100, http://ruccs.rutgers.edu./roa.html.

Grimshaw, J. & Samek-Lodovici, V. (1995). Optimal subjects. In J. Beckman, L. W. Dickey, & S. Urbanczyk (Eds.), *Papers in Optimality Theory*, number 18 in University of Massachusetts Occasional Papers in Linguistics (pp. 589–606). Amherst, MA: Graduate Linguistic Student Association.

Harman, G. (1980). Two quibbles about analyticity and psychological reality. *The Behavioral and Brain Sciences*, *3*, 21–2.

Jakobson, R. (1968). *Child language : aphasia and phonological universals*. Janua linguarum. The Hague: Mouton.

Kent, R. D. (1992a). The biology of phonological development. In C. A. Ferguson, L. Menn, & C. Stoel-Gammon (Eds.), *Phonological Development: models, research, implications* (pp. 65–90). Timonium MD: York Press.

Kent, R. D. (1992b). Phonological Development as Biology and Behavior. In R. S. Chapman (Ed.), *Process in Language Acquisition and Disorders* (pp. 67–85). St Louis, MS: Mosby-Year Book, Inc.

Kent, R. D. & Miolo, G. (1995). Phonetic Abilities in the First Year of Life. In P. Fletcher & B. MacWhinney (Eds.), *The Handbook of Child Language*. Blackwell.

Kiparsky, P. & Menn, L. (1977). On the Acquisition of Phonology. In J. Macnamara (Ed.), *Language Learning and Thought*. New York: Academic Press.

Koskenniemi, K. (1983). Two-Level Model for Morphological Analysis. In *Proceedings of the Eighth International Joint Conference on Artificial Intelligence*, (pp. 683–685).

Legendre, G., Raymond, W., & Smolensky, P. (1993). Optimality-Theoretic typology of case and grammatical voice systems. In *Proceedings of the 19th Meeting of the Berkeley Lingustics Society*. Rutgers Optimality Archive, ROA-3, http://ruccs.rutgers.edu./roa.html.

Li, M. & Vitanyi, P. M. B. (1989). Inductive Reasoning and Kolmogorov Complexity. In *Proceedings of the Fourth Annual Conference on Structure in Complex ity Theory*, (pp. 165–185).

Li, M. & Vitanyi, P. M. B. (1993). *An introduction to Kolmogorov complexity and its applications*. Texts and Monographs in Computer Science. Springer Verlag.

Matthei, E. (1989). Crossing boundaries: more evidence for phonological constraints on early multi-word utterances. *Journal of Child Language*, *16*, 41–54.

McCarthy, J. (1996). Extensions of faithfulness: Rotuman revisited. Rutgers Optimality Archive, ROA-110, http://ruccs.rutgers.edu./roa.html.

McCarthy, J. & Prince, A. (1995). Faithfulness & reduplicative identity. In J. Beckman, L. W. Dickey, & S. Urbanczyk (Eds.), *Papers in Optimality Theory*, number 18 in University of Massachusetts Occasional Papers in Linguistics (pp. 249–384). Amherst, MA: Graduate Linguistic Student Association. Rutgers Optimality Archive, ROA-60, http://ruccs.rutgers.edu./roa.html.

Menn, L. (1983). Development of Articulatory, Phonetic and Phonological Capabilities. In *Language Production*, volume 2 (pp. 3–50). London: Academic Press.

Newton, I. (1953). *Newton's Philosophy of Nature: Selections from His Writings*. New York: Haffner.

Orgun, C. O. (1994). Monotonic cyclicity and Optimality Theory. In Gonzalez, M. (Ed.), *Proceedings of the 24th Annual Meeting of the North-East Linguistic Society*, (pp. 461–474)., Amherst, MA. Graduate Linguistic Student Association. Rutgers Optimality Archive, ROA-123, http://ruccs.rutgers.edu./roa.html.

Orgun, C. O. (1995). Correspondence and identity constraints in two-level OT. Rutgers Optimality Archive, ROA-62, http://ruccs.rutgers.edu./roa.html.

Pinker, S. (1989). *Learnability and Cognition: the acquisition of argument structure*. Cambridge, MA: MIT/Bradford.

Pinker, S. (1994). *The Language Instinct*. Penguin Books.

Popper, K. (1959). *The Logic of Scientific Discovery*. London: Hutchinson.

Prince, A. & Smolensky, P. (1993). Optimality theory: constraint interaction in generative grammar. Technical Report TR-2, Rutgers University, Center for Cognitive Science. Same as Prince and Smolensky (to appear).

Prince, A. & Smolensky, P. (to appear). *Optimality theory: constraint interaction in generative grammar*. Cambridge, MA: MIT Press. Same as Prince and Smolensky (1993).

Quartz, S. R. & Sejnowski, T. J. (in press). The Neural Basis of Cognitive Development: A Constructivist Manifesto. *The Behavioral and Brain Sciences*, *??*, ??

Reichenbach, H. (1956). *The Direction of Time*. Berkeley: University of California Press.

Rissanen, J. (1978). Modelling by Shortest Data Description. *Automatica*, *14*, 445–71.

Rissanen, J. (1982). Estimation of Structure by Minimum Description Length. *Circuit systems signal Processing*, *1*, 395–396.

Rissanen, J. (1987). Statistical Complexity. *Journal of the Royal Statistical Society*, *B49*, 223–39,252–65.

Russell, K. (1995). Morphemes and Candidates in Optimality Theory. Rutgers Optimality Archive, ROA-44, http://ruccs.rutgers.edu./roa.html.

Salmon, W. (1975). Theoretical Explanation. In S. Korner (Ed.), *Explanation* (pp. 118–145). Blackwell.

Salmon, W. (1978). Why ask "why"? *Proceedings and Addresses of the Americal Philosophical Association*, *51*, 683–705.

Salmon, W. (1984). *Scientific Explanation and the Causal Structure of the World*. Princeton, NJ: Princeton University Press.

Schwartz, R. & Leonard, L. (1982). Do Children Pick and Choose: An Examination of Phonological Selection and Avoidance in Early Lexical Acquisition. *Journal of Child Languge*, (9), 319–336.

Scobbie, J. M. (1991). *Attribute Value Phonology*. PhD thesis, University of Edinburgh.

Scobbie, J. M. (1997). *Autosegmental Representation in a Declarative Constraint-Based Framework*. New York NY: Garland.

Scobbie, J. M., Coleman, J., & Bird, S. (1996). Key aspects of declarative phonology. In J. Durand & B. Laks (Eds.), *Current Trends in Phonology: models and methods*, volume 2 (pp. 685–710). European Studies Research Institute, University of Salford Publications.

Sells, P., Rickford, J., & Wasow, T. (1994). An Optimality-Theoretic approach to variation in negative inversion in AAVE. Stanford University. Rutgers Optimality Archive, ROA-53, http://ruccs.rutgers.edu./roa.html.

Smolensky, P. (1996). The Initial State and 'Richness of the Base' in Optimality Theory. Technical Report JHU-CogSci-96-4, Department of Cognitive Science, Johns Hopkins University. Rutgers Optimality Archive, ROA-154, http://ruccs.rutgers.edu./roa.html.

Sober, E. (1988). *Reconstructing the Past*. Cambridge MA: MIT Press.

Sokolov, J. & Snow, C. (1994). The changing role of negative evidence in theories of language acquisition. In *Input and Interaction in Language Acquisition* (pp. 38–55). London: Cambridge University Press.

Speas, M. (1995). Generalized control and null objects in Optimality Theory. In J. Beckman, L. W. Dickey, & S. Urbanczyk (Eds.), *Papers in Optimality Theory*, number 18 in University of Massachusetts Occasional Papers in Linguistics (pp. 637–654). Amherst, MA: Graduate Linguistic Student Association.

Stampe, D. (1979). *A Dissertation on Natural Phonology*. New York: Garland.

Tesar, B. & Smolensky, P. (1993). The Learnability of Optimality Theory: an Algorithm and some Basic Complexity Results. Technical report, Department of Computer Science, University of Colorado, Boulder. Rutgers Optimality Archive, ROA-2, http://ruccs.rutgers.edu./roa.html.

Tesar, B. & Smolensky, P. (1996). Learnability in Optimality Theory (long version). Technical report, Department of Cognitive Science, Johns Hopkins University and Center for Cognitive Science, Rutgers University. Rutgers Optimality Archive, ROA-156, http://ruccs.rutgers.edu./roa.html.

Tesar, B. B. (1995). *Computational Optimality Theory.* PhD thesis, University of Colorado.

Valian, V. (1990). Logical and psychological constraints on the acquisition of syntax. In L. Frazier & J. de Villiers (Eds.), *Language Processing and Language Acquisition* (pp. 119–145). Dordrecht: Kluwer.

Vihman, M. M. (1992). Early Syllables and the Construction of Phonology. In C. A. Ferguson, L. Menn, & C. Stoel-Gammon (Eds.), *Phonological Development: models, research, implications* (pp. 393–422). Timonium MD: York Press.

Wallace, C. & Freeman, P. (1992). Single-factor analysis by minimum message length estimation. *Journal of the Royal Statistical Society*, *B54*, 195–209.

Wallace, C. S. & Boulton, D. M. (1968). An information measure for classification. *Computer Journal*, *11*, 185–195.

Woolford, E. (1995). Object agreement in Palauan: Specificity, humanness, economy, and optimality. In J. Beckman, L. W. Dickey, & S. Urbanczyk (Eds.), *Papers in Optimality Theory*, number 18 in University of Massachusetts Occasional Papers in Linguistics (pp. 655–702). Amherst, MA: Graduate Linguistic Student Association. Rutgers Optimality Archive, ROA-55, http://ruccs.rutgers.edu./roa.html.