Markedness in phonology and in syntax:
the problem of grounding[*]

Ricardo Bermúdez-Otero[a] & Kersti Börjars[b]

[a] University of Newcastle upon Tyne, [b] University of Manchester

r.bermudez-Otero@ncl.ac.uk
phone: +44 (0)191 222 3519
fax: +44 (0)191 222 8708

k.borjars@man.ac.uk
phone: +44 (0)161 275 3042
fax: +44 (0)161 275 3187

ABSTRACT

This article adopts the perspective of Optimality Theory (OT) to address the question whether phonology and syntax are equally autonomous. We show that OT enjoys the same advantages and encounters the same problems in syntax as in phonology; this suggests that markedness plays an equally important rôle in both components of language. Most markedness constraints, however, are clearly grounded: although they refer to specifically linguistic categories (self-containment), they typically display some degree of functional adaptation to the demands of performance (nonarbitrariness). In consequence, phonology and syntax should be expected to be grounded to a similar degree. *Pace* Hale & Reiss (2000), however, the postulation of grounded markedness constraints in the theory of grammar does not violate Ockham's Razor. In particular, we show that markedness cannot be equated with performance difficulty, and we demonstrate that infants require knowledge of markedness during language acquisition in order to transcend the limitations of inductive generalization. However, this does not necessarily imply that knowledge of markedness is innate; we argue, rather, that most markedness constraints may in fact emerge in the course of linguistic development through the child's monitoring of her own performance.

KEYWORDS: Optimality Theory; markedness; autonomy; grounding; innateness; phonology; syntax

1. Autonomy in syntax and in phonology

Ever since the advent of generative theory, research into language has been overshadowed by the dispute between formalist and functionalist approaches. At stake in this controversy is the *autonomy* thesis: i.e. the hypothesis that linguistic knowledge is independent from, and irreducible to, the facts of language use. The issues involved are notoriously complex. Following Croft (1995), for example, it is useful to consider the autonomy thesis as encompassing two separate claims: grammatical patterns are *arbitrary* insofar as they fail to match external functions; in addition, grammar is *self-contained* inasmuch as it manipulates specifically linguistic categories and does not refer to extragrammatical entities. The complexity of the overall problem is reflected in a continuum of opinion ranging from radical formalism to radical functionalism.

Historically, the autonomy debate has raged with equal fury in syntax and in phonology. Thus the contrast between the extreme formalism of Minimalist syntax (Chomsky, 1995; see below) and the extreme functionalism of, say, García (1979) and Kalmár (1979)[1] finds parallels in phonological research: e.g. whereas Hale & Reiss (2000) argue that phonetic substance plays absolutely no rôle in phonology, some versions of Laboratory Phonology dismiss the distinction between phonetics and phonology as altogether untenable (Ohala, 1990; Pierrehumbert *et al.*, 2000). With the rise of the Minimalist Program, however, mainstream generative syntax has witnessed a radicalization of the autonomy thesis. For Minimalists, syntax is optimally designed to meet the requirements that hold at its interface with other mind-internal systems. However, Minimalists emphasize that, by being a 'perfect' solution to interface conditions, syntax is in no sense rendered well designed for use (Chomsky *et al.*, 2002: 106-7); in this sense, it is radically arbitrary. In addition, Minimalists claim that the syntactic computational system is universally invariant: parametrization no longer resides in syntax itself, but rather in the choice of feature specifications for lexical items (including functional categories).[2]

The radically autonomous vision of syntax that is currently being pursued in Chomskyan circles has lent new urgency to the question whether phonology and syntax can in fact possess autonomy to the same degree (Bermúdez-Otero, 2002: 404). This problem has a long history (e.g. S. Anderson, 1981), but it has lately received new and radical answers, notably through the Representational Hypothesis advanced by Burton-Roberts (2000), Carr (2000, this volume), and Burton-Roberts & Poole (this volume). In the Representational Hypothesis, the language faculty enjoys all the attributes of autonomy, in line with Minimalist thought; phonology, however, stands outside UG and consists of a set of cultural conventions for the external representation of internal syntactic-semantic objects in the phonetic medium. From this position, Carr (2000, this volume) claims that the acquisition of phonology – or, at least, of what Burton-Roberts & Poole call 'narrow phonology', i.e. prosody and melody – calls for a thoroughly empiricist (i.e. functionalist) approach; see also Burton-Roberts (2000: 61).

In this paper we wish to challenge this dichotomy between autonomous syntax and functional phonology. However, we shall not follow in the tradition of those linguists who argue for similarities between the representational primitives and structural relations found in both dimensions of language. Such views have been expounded in different guises by J. Anderson (1986, this volume), J. Anderson & Ewen (1987), Pierrehumbert (1990), and van der Hulst (2000, this volume), among others; and they underpin the account of language evolution given by Carstairs-McCarthy (1999), criticized by Tallerman (this volume). Though we will not explicitly argue against such proposals (see Carr, this volume), we view them with profound skepticism: to our minds there are fundamental differences in nature between syntactic and phonological primitives and relations. To a large extent, these

differences may follow from the dual articulation of language: the elements of syntax have meaning; those of phonology do not.

Indeed, representational primitives and structural relations can be expected to be identical in phonology and in syntax only if one holds extreme formalist or extreme functionalist views of both. In the first case (that of someone for whom both phonology and syntax are strictly autonomous), there is no reason why representations in both components of language could not be built out of the same purely formal elements, as these would stand in a purely arbitrary relationship with linguistic substance. This, in essence, is van der Hulst's (2000, this volume) position. For the extreme functionalist, who views all language acquisition as a case of purely general learning, syntactic and phonological representations could be identical in the sense of both consisting of – say – clusters of exemplars (Johnson, 1997; Pisoni, 1997).

In this paper, however, we take a less extreme view. Adopting the perspective of Optimality Theory (henceforth OT), we argue that key aspects of both phonological and syntactic knowledge are grammatically encoded in terms of the language-specific ranking of universal violable markedness constraints. The knowledge so encoded is *self-contained* but partially *nonarbitrary*: it refers to specifically linguistic (syntactic or phonological) categories, but displays some degree of functional adaptation to the requirements of performance (§2). We use the term 'grounding' to refer to this state of affairs; in our view, both phonology and syntax are *grounded*. In support of this position we adduce evidence that markedness constraints play similar rôles in phonology and in syntax (§3). We show, in particular, that the deployment of OT leads to the same advantages and difficulties in syntax as it does in phonology (§4, §5): contrary to a common opinion, for example, the problem of ineffability is as pressing in phonology as in syntax (cf. Pesetsky, 1997, 1998).

The concept of grounding is, however, highly controversial. In §6, therefore, we face up to the charge of 'substance abuse' that has lately been levelled against OT (Hale & Reiss, 2000). We demonstrate that the postulation of grounded markedness constraints in the theory of grammar does not violate Ockham's Razor in the way that Hale & Reiss suggest. In particular, we show that markedness cannot be reduced to performance: there is, in fact, no consistent connection between linguistic markedness relationships and performance difficulty landscapes. We argue, moreover, that markedness relationships do not enter language simply by induction over primary linguistic data subject to performance bias; on the contrary, we provide empirical evidence that, during acquisition, infants require knowledge of markedness in order to transcend the limitations of inductive generalization. Having thus established the need for markedness constraints, in §7 we explore issues related to their origin: we show that not all markedness constraints need be supplied innately by UG, and we tentatively suggest that most constraints may in fact emerge in the course of linguistic development.

To some extent, then, our paper can be read as a response to the radically formalist views expressed by Hale & Reiss (2000) and by Burton-Roberts & Poole (this volume). It may therefore be interesting at this point to compare our general strategy with that of Burton-Roberts & Poole, as their premises and methodology are so fundamentally different from ours. Our approach here is to focus on what clearly are two different domains: phonology and syntax as traditionally understood.[3] We compare these two domains with respect to empirical attributes related to markedness: conspiracies, the emergence of the unmarked, ineffability, and so forth. Our comparison reveals close parallels between the two domains, and from this we conclude that phonology and syntax are, in the relevant respects, similar.

Burton-Roberts & Poole, in contrast, define the properties of the language faculty aprioristically on the basis of an ultra-autonomist agenda and proceed to squeeze out of Language whatever does not fit into this conceptual design, redefining the nature and scope of phonology in the process. Naturally, phonology falls outside UG as they define it – but,

more startlingly, so does most of what we know as syntax.[4] The fact is wittily highlighted by van der Hulst (this volume), who points out that what Burton-Roberts & Poole call Language is regarded by Jackendoff (2002) as outside Language proper, whereas that which Burton-Roberts & Poole exclude from Language (namely, all phonology and most of what is usually regarded as syntax) is precisely what Jackendoff calls Language. This amounts to much more than a terminological squabble: Chomsky advanced the innateness hypothesis and the concept of UG as a solution to Plato's Problem in respect of linguistic knowledge (see e.g. Chomsky, 1986); since Burton-Roberts & Poole exclude all of phonology and most of traditional syntax from the realm of UG, one is entitled to ask whether Plato's Problem does not arise in these domains and, if it does, how it should be addressed.[5]

2. The view from OT

When one attempts to ascertain the implications of OT for the autonomy of linguistic knowledge, several considerations suggest that answers will emerge most readily from applications of the theory to phonology. Foundational work in OT (notably Prince & Smolensky, 1993) addressed itself almost exclusively to phonological problems, and so it is in phonology that OT has been practised the longest; similarly, OT has rapidly become the dominant paradigm in phonology, whereas it continues to face tough competition in syntax. In consequence, it is in phonology that OT has produced the largest volume of research and achieved the highest number of results. Thanks to well-established scientific disciplines such as phonetics, moreover, the nonlinguistic domains with which phonology interacts are comparatively well-understood, so that phonologists can benefit from this improved understanding when tackling the issue of autonomy.

Even within phonology, however, OT has developed in different directions, with the result that researchers often make significantly different assumptions on issues that have a bearing on the autonomy debate. Notably, whilst it has been common practice to assume that the universal constraint set (CON) is supplied innately, some researchers have suggested that learners discover constraints in the course of language acquisition: Hayes (1999a), for example, suggests that children rely on an innately endowed algorithm to identify constraints on the basis of their experience of phonetic difficulty (see §7.2 below). Similarly, mainstream work in OT phonology assumes that constraints are exclusively formulated in terms of specifically phonological categories such as features, segments, syllables, and feet. However, a sizeable body of work endorses the view that constraints should be granted direct access to phonetic properties, even when these are universally noncontrastive (e.g. Kirchner, 1997; Steriade, 1997, 2000; Zhang, 2000); see §6.3 below. For the sake of argument, we shall here explore the implications of the mainstream approach and address alternative positions later.[6]

To begin with, it is interesting to note that OT does not allow the strict separation of the natural from the conventional that is implicit in the Minimalist Program and is made explicit in the Representational Hypothesis. In this area, OT supports a conservative position, closer in fact to Government and Binding theory (GB) than to Minimalism, in which language-particular grammars involve a mixture of the biological and the conventional. The space of possible grammars is defined biologically, and indeed the substance of language-particular grammars (i.e. the constraints) is biologically determined; but the ranking of constraints – like the setting of parameters in GB – is ultimately a matter of convention. This is true even in modified versions of OT such as Hayes's: although in his theory constraints are not supplied innately by UG, their formulation is nonetheless altogether determined by biological givens, namely the innate constraint-discovery algorithm plus the physical and physiological facts that cause phonetic difficulty; in this sense, the membership of CON itself remains beyond the bounds of convention.

An equally nuanced approach is suggested in relation to the questions of arbitrariness and self-containment (Croft, 1995). In mainstream versions of OT, phonology is *self-contained*: phonological constraints are formulated in terms of specifically phonological categories, rather than phonetic events or properties. Take for example CODACOND-Place, the constraint that triggers the neutralization of consonantal place contrasts in the syllable coda (Itô, 1989; Kager, 1999: 131; McCarthy, 2002: 145):

(1)   CODACOND-Place

        Assign one violation mark for every C-Place node that is exhaustively dominated by rhymal segments.

As given in (1), CODACOND-Place refers to a node in feature geometry (C-Place) and to a position in syllable structure (the rhyme). These categories are cognitive and, as such, ontologically distinct from phonetic entities such as articulatory gestures and acoustic spectra. Nor does the relationship between phonological categories and phonetic events involve simple one-to-one correspondence: notably, it is impossible to provide an articulatory or acoustic definition of the syllable (Ladefoged & Maddieson, 1996: 281-2).[7]

However, whereas phonological constraints possess the attribute of self-containment, there is an important sense in which they are not *arbitrary*. The example of CODACOND-Place is again representative. Phonetically, the place cues of a consonant are relatively difficult to discriminate if that consonant is unreleased, i.e. not followed by a more open articulation (see e.g. Myers, 1997: 133). Since coda consonants are typically – though, crucially, not necessarily – realized without release, it follows that, perceptually, place contrasts will in general be difficult to maintain in the coda. Accordingly, the constraint CODACOND-Place is phonetically motivated, in that it enforces phonological place neutralization in environments where the phonetic realization of contrast is problematic. Following Archangeli & Pulleyblank (1994), constraints manifesting these properties are usually described as *grounded*. In the light of our discussion, we shall henceforth use the term 'grounding' specifically to refer to the status of an item of linguistic knowledge that exhibits self-containment but nonarbitrariness.

Admittedly, the basic principles of OT do not logically require constraints to be grounded. Nonetheless, grounding asserts itself as an overwhelming empirical fact, in phonology at least. Time and again, the analysis of patterns of distribution and alternation, both language-internally and crosslinguistically, has led to the postulation of phonological constraints which, upon examination, have turned out to be phonetically nonarbitrary – so much so that grounding is often implicitly regarded as a hallmark of genuine or 'principled' constraints (Hayes, 1999a: §3). We conclude, therefore, that, insofar as the phonological applications of the theory are representative, OT suggests that linguistic knowledge is neither radically autonomous nor purely functional; rather, it is self-contained but partially nonarbitrary, i.e. it is grounded.

3. OT and syntax
Of course, the claim that linguistic knowledge is to a large extent grounded raises thorny conceptual problems, eloquently highlighted in Hale & Reiss (2000). We will however put off the discussion of these issues until §6 and §7, and now address the question whether, like phonology, syntax is grounded. In this connection, an obvious strategy to pursue would be to examine the syntactic constraints postulated so far in the optimality-theoretic literature and to

determine whether they exhibit self-containment but nonarbitrariness. However, this strategy proves impractical for a number of reasons.

First, there is relatively little consensus as to how OT syntax works. One difficulty concerns the status of OT as a metatheory: the basic postulates of OT are compatible with a wide range of approaches to the nature of syntactic representations; as a result, assumptions vary as regards both the operation of the candidate generator (GEN) and the membership of the candidate set. There is arguably some OT work in syntax that can be described as relatively neutral with respect to theoretical assumptions (e.g. Börjars & Donohue, 2000; Aissen, 2001). Similarly, many of the constraints postulated in more theory-specific research represent widely accepted generalizations and would therefore be used in roughly similar forms in all kinds of work on OT syntax; a likely example is the constraint OP-SPEC, which captures the fact that operators such as *wh*-phrases tend to occupy peripheral positions. As regards the nature of GEN, however, there is a fundamental split between transformational and parallel-correspondence approaches. In the first view, GEN is a transformational engine in which chains link elements at different stages of a transformational derivation or calculation. Within this approach we can distinguish work allied to GB (GB-OT: e.g. Legendre *et al.*, 1995; Grimshaw, 1997; Legendre, Smolensky & Wilson, 1998) and work allied to Minimalism (Minimalism-OT: e.g. Speas, 1997, 2001; Müller, 2001). An alternative view is represented by research combining Lexical Functional Grammar and OT (LFG-OT: e.g. Bresnan, 2000a, 2001a; Sells, 2001a, b). In LFG-OT, GEN encapsulates a theory of parallel correspondence where candidates consist of coindexed dimensions of linguistic information: e.g. c-structure and f-structure. These different conceptions of GEN entail significant differences in the formulation of constraints,[8] which are likely to affect judgements about their grounding. An explicit comparison of transformational and parallel-correspondence approaches can be found in Bresnan (2000a).

To compound this difficulty, the nonlinguistic domains with which syntax interfaces are relatively poorly understood, and so it proves difficult to judge the extent to which syntactic constraints are arbitrary or motivated.[9] Recent research has nonetheless made it clear that the notion of grounding is relevant not only to the 'aspects of language closest to the mouth and ear' (Pesetsky, 1997: 135), but also to more abstract cognitive capacities. In this connection one should not ignore the wealth of ideas developed within Cognitive Grammar and other functionalist approaches (see e.g. Haiman, 1983; Givón, 1995; Croft, 2001). For an example of work in OT syntax using constraints that are clearly grounded, see Bresnan (2001b); Bresnan & Aissen (2002) provide a general discussion of functionality and OT constraints.

Faced with such difficulties, one would not get very far by examining individual syntactic markedness constraints and attempting to decide whether or not they are grounded. Selecting this or that constraint would beg as many questions as it would answer, and one would be unlikely to reach conclusions that would command assent beyond the framework in which one's chosen examples were formulated. In consequence, we have opted to probe the nature of syntax using a more indirect strategy – but ultimately a more ambitious one.

First, we note that optimality-theoretic constraints display a set of properties which bespeak a similar nature and origin. Most evidently, they all share the nontrivial attributes of universality and of minimal violability under strict ranking. In addition, they interact promiscuously and on an equal footing: within the same hierarchy, any constraint may conflict with any other. Since, therefore, constraints are so similar to each other and interact so closely, it seems implausible to suggest that they could have widely different phylogenetic or ontogenetic origins. Our discussion in §2, however, showed that phonological constraints are grounded. Accordingly, if syntactic constraints, being similar in nature and behaviour, have similar origins, we would expect them also to be grounded.

The key question, then, is not whether phonological and syntactic constraints are similarly grounded; to the extent that it can be posed, this question should be answered, we suggest, in the affirmative. Rather, one needs to address a more fundamental problem: is syntax, like phonology, governed by universal violable markedness constraints? To tackle this deeper issue, §4 and §5 examine the advantages and problems attendant on applying OT to syntax. We ask, in other words, whether the empirical successes and failures of OT are specific to phonology or generalize to syntax. Note, however, that not all the results of OT stem from its reliance on markedness constraints: the notorious challenges that phonological opacity poses for the theory, for example, have little to do with the use of markedness constraints, but rather relate to OT's rejection of serialism in favour of a parallel architecture (see e.g. Bermúdez-Otero, 1999; McCarthy, 2002: §3.3). Therefore, since our ultimate concern is with grounding, we shall specifically focus on the consequences of asserting that syntactic systems maximize harmony with respect to ranked sets of universal violable markedness constraints, and we will accordingly de-emphasize other aspects of OT, such as globality and parallelism, which do not seem so directly implicated in grounding. In sum, insofar as relying on markedness constraints leads syntacticians to the same achievements and perplexities as phonologists, phonology and syntax may be expected to be similarly grounded and, therefore, to possess or lack autonomy to a similar degree.

Interestingly, there is an extremely wide range of opinion as to how much of syntax is optimality-theoretic; for a survey, see Boersma, Dekkers & van de Weijer (2000: §C.1).[10] Pesetsky (1997, 1998), for example, argues that violable constraints only govern certain areas of syntax, notably those related to what he calls 'pronunciation' (i.e., roughly, the overt phonological realization of syntactic elements). In his view, other aspects of syntax – what he calls 'movement' – cannot be subject to optimality-theoretic constraints. His basic claim is that movement works in 'clash-and-crash' fashion: a syntactic expression cannot be grammatical unless it satisfies all constraints on movement simultaneously. According to Pesetsky, therefore, movement cannot be handled within OT, where it raises the problem of ineffability; see §5.1 below.

There are several problems with this line of argument.[11] First, many of the 'pronunciation' issues that Pesetsky addresses relate to the phonological realization of chains, whose existence Pesetsky presents as an incontrovertible fact. Chains can however be regarded as mere artifacts of a particular approach to syntax; what is beyond question is, rather, the existence of long-distance dependencies such as may hold, for instance, between a verb and an argument realized as a sentence initial *wh*-constituent. These long-distance dependencies need not involve chains unless one conflates the representation of constituency with that of grammatical relations by assuming that certain semantic properties are associated with particular structural positions. A logical alternative is to conceive of different types of information as expressed on parallel planes of representation, like c-structure and f-structure in LFG, subject to imperfect correspondence (Bresnan, 2001a; see also Chomsky *et al.*, 2002: 120; Jackendoff , 2002). Note that, typologically, grammatical relations are often signalled without recourse to constituency. In addition, transformational theories of syntax often postulate chains that are not motivated by genuine long-distance dependencies, but arise from attempts to handle phenomena such as inflection through syntactic movement. The key point, then, is that Pesetsky attempts to restrict the application of OT to syntax on the basis of a distinction between movement and pronunciation which, instead of reflecting incontrovertible empirical facts, is actually highly theory-internal. Quite simply, in a parallel-correspondence framework a 'pronunciation' constraint such as SILENTTRACE ("Don't pronounce the traces of a moved constituent"; Pesetsky, 1997: 153) makes no sense; see Bresnan (2000a).

Secondly, Pesetsky's proposals predict that the distinction between movement and pronunciation is closely correlated with a contrast between clash-and-crash and do-your-best

behaviours: clash-and-crash effects allegedly show that movement calls for an analysis in terms of inviolable principles and parameters, whereas do-your-best phenomena show pronunciation to be governed by soft constraints. In §5.1 below, however, we show that ineffability effects are found throughout the grammar, including phonology; ineffability therefore fails to provide empirical support for limiting the application of OT to a particular area of syntax. In addition, the same syntactic conflict may lead to a compromise ('do your best') in one language and to ineffability ('crash') in another; see Börjars & Vincent (2000) for some examples. If, as Pesetsky suggests, the contrast between clash-and-crash and do-your-best behaviours serves as an empirical diagnostic for the boundary between movement and pronunciation, then this crosslinguistic variation in the resolution of syntactic conflicts leads to the unpalatable conclusion that the division of labour between movement and pronunciation is largely language-specific.

Pesetsky also appeals to gradient grammaticality judgements to identify syntactic phenomena that resist analysis in OT terms. He argues that OT cannot deal with gradience because in OT "there are no silver or bronze medals" (Pesetsky, 1997: 169). In this connection, we will merely observe that in Principles-and-Parameters syntax the notion of coming second best is equally problematic. These phenomena will therefore cause headaches to most syntacticians, regardless of theoretical allegiance.

All in all, then, we are not persuaded by Pesetsky's arguments that violable markedness constraints only govern certain areas of syntax. Rather, in the light of the evidence surveyed in §4 and §5, we will take the view that, at present, there is every reason to suppose that the interaction of violable constraints is responsible for all syntactic phenomena.[12]

4. Empirical advantages of OT

In this section we consider whether the application of OT yields the same empirical benefits in syntax as it does in phonology. Assuming that markedness constraints constitute a locus of grounding (§3), we shall specifically concentrate on those advantages of the theory that derive from the use of universal markedness constraints violable under strict ranking. These advantages include providing a straightforward account of conspiracies (§4.1), getting rid of *ad hoc* hedges in the formulation of grammatical statements (§4.2), and predicting emergence-of-the-unmarked effects (§4.3). Our discussion will show that these features of OT are as desirable in syntax as they are acknowledged to be in phonology. We suspect that other advantages of OT, such as accounting for top-down and chicken-and-egg effects (McCarthy, 2002: §3.3.2.3, §3.3.2.4), also cut across the phonology/syntax divide, but these results are more closely associated with other aspects of the theory, such as parallel constraint evaluation.

4.1. Conspiracies

For phonologists, one of the most attractive features of OT is its ability to predict the existence of conspiracies (see e.g. Kager, 1999: §2.1.1.2; Davis, 2000; McCarthy, 2002: §3.1.4.3). A conspiracy is said to occur when several formally diverse grammatical processes serve a common output target. OT's reliance on markedness constraints predicts that such states of affairs will arise because markedness constraints define conditions of maximal harmony for outputs, but do not specify how such conditions are to be fulfilled. When a choice of possible repair strategies exists, the outcome is determined independently by the ranking of other constraints. Moreover, since these other constraints may be sensitive to configurations present only in a subset of the relevant marked structures, different repairs may be appropriate in different local contexts. In this sense, OT predicts the existence of both *crosslinguistic conspiracies* (where different languages resort to different strategies to avoid

the same marked structure) and *language-internal conspiracies* (where a language deploys different repair processes for the same marked configuration in different contexts).

Pater (1999) provides a celebrated example of a crosslinguistic conspiracy. The display in (2), adapted from Kager (1999: 87), lists the processes that different languages resort to in order to avoid marked sequences of a nasal consonant followed by a voiceless obstruent, which violate the constraint *NC[-voi].[13]

(2) **The *NC[-voi] conspiracy**

	Strategy	*Rule*	*Language*
a.	Nasal substitution	[−son, −voi] → Ø / [+nas] __	Indonesian
b.	Nasal deletion	[+nas] → Ø / __ [−son, −voi]	Kelantan
c.	Post-nasal voicing	[−son] → [+voi] / [+nas] __	Puyo Pungo
d.	Denasalization	[+nas] → [−nas] / __ [−son, −voi]	Mandar

Davis (2000: 122) gives a striking example of a language-internal conspiracy: in Korean, no fewer than four separate rules conspire to enforce CONTACT, the markedness constraint requiring that sonority should not rise across a syllable boundary (Vennemann, 1988; Clements, 1990, 1992; Bat-El, 1996).

(3) **The CONTACT conspiracy in Korean**

	Process	*Example*		
a.	anticipatory nasalization	/kuk+mul/	→	[kuŋ.mul]	'broth'
b.	anticipatory lateralization	/non-li/	→	[nol.li]	'logic'
		/tikɨt+liɨl/	→	[ti.kɨl.li.ɨl]	'*t* and *l*'
c.	perseverative nasalization	/kam-li/	→	[kam.ni]	'supervision'
d.	nonassimilatory nasalization	/pəp-li/	→	[pəm.ni]	'principle of law'

OT's ability to account for conspiracies proves to be as advantageous in syntax as it is in phonology. Like phonological ones, syntactic markedness constraints give rise to both crosslinguistic and language-internal conspiracies. An example of the former arises in connection with the requirement that subjects should be overtly marked; in (4a) we formulate this requirement as a correspondence constraint in the framework of LFG (see Bresnan, 2000a: 349-50). As it turns out, different languages deploy diverse strategies for satisfying MARK-SUBJ, as listed in (4b); for head-marking and dependent-marking, see Nichols (1986).[14]

(4) **The subject marking conspiracy**

a. MARK-SUBJ
Overtly mark f-structure SUBJ in c-structure.

	Strategy	*Example*
	Structural position (word order)	English
	Head-marking (on verb)	Abkhaz, Tzutujil
	Dependent-marking (on NP)	Chechen, Dyirbal

It is important to note that MARK-SUBJ is a genuinely violable markedness constraint: e.g. it fails to be satisfied in the Swedish sentence in (5), which is rendered ambiguous by the requirement that the verb should be in second position.[15]

(5) Oscar gillar Sara.
     Oscar like.FIN Sara
     'Oscar likes Sara' or 'Sara likes Oscar[TOPIC]'

In languages such as German, verb-second requirements also give rise to language-internal conspiracies. Assume a highly ranked constraint requiring the finite verb to occur in second position – or rather two separate constraints creating this effect (S. Anderson, 1996, 2000; Legendre, 2000):

(6) **Constraints responsible for the V2 effect in German**

NONINITIAL(V_{fin}) » ALIGN(Left, V_{fin}; Left, Clause)[16]

The constraints in (6) can be satisfied in a number of different ways: e.g. through canonical $SV_{fin}O$ order, or by placing a topicalized element before the finite verb to give $O_{TOP}V_{fin}SAdv$ or $Adv_{TOP}V_{fin}SO$. However, data from impersonal passives provides particularly strong support for a conspiratorial analysis:

(7) a. Gestern wurde   getanzt.
         yesterday PASS.PAST dance.PART
     b. Es   wurde   gestern   getanzt.
         EXPL   PASS.PAST yesterday dance.PART
         'People danced yesterday.'
     c. *Gestern wurde   es   getanzt.
          yesterday PASS.PAST EXPL dance.PART

The grammaticality of (7a) and the ungrammaticality of (7c) demonstrate that the expletive in (7b) is not there to satisfy a requirement that the clause should have a subject, along the lines of the Extended Projection Principle of GB. Rather, the insertion of an expletive in (7b) constitutes yet another strategy in a conspiracy to satisfy a linearization constraint involving the finite verb.[17] As Legendre (2001: 4–11) points out, data like (7) clearly illustrate the drawbacks of analyses that account for the presence of expletive pronouns by invoking inviolable parameter settings.

From a transformational viewpoint (see §3 again), the constraint OB-HD provides another example of a syntactic conspiracy. OB-HD (for 'OBLIGATORYHEADS'; see e.g. Kager, 1999: 349) requires the heads of phrases to be filled. In English this may involve a lexical

head generated in the phrase (e.g. a modal verb in I), a moved head (e.g. a lexical verb raised to I), or a dummy element (e.g. *do* in I); see Vikner (2001).

4.2. Removing hedges

Another advantage of OT comes to light in cases where an optimality-theoretic constraint corresponds fairly directly to some other statement (such as an inviolable principle, a parameter, or a rule) in a different grammatical framework. Typically, the constraint proves formally simpler, in that it dispenses with *ad hoc* hedges or exception clauses incorporated into the alternative statement. This is because, in frameworks other than OT, grammatical statements are required to hold strictly true in their relevant domain; in consequence, systematic patterns of exception must be written into their formulation. In OT, in contrast, constraint violation results from domination under strict ranking; accordingly, the set of circumstances where a constraint may not hold true falls out from factorial typology. There thus turns out to be a close correlation between violability and formal simplicity.

In phonology, the theory of extrametricality provides a striking example of this correlation. In a parametric framework (e.g. Hayes, 1995), the extrametricality parameter has to be supplemented with two hedges:

(8)   **Hedges on extrametricality in a parametric framework**

  a.   *Nonexhaustivity* (Hayes, 1995: 58)
         Extrametricality is blocked if the entire domain of stress assignment would otherwise be extrametrical.

  b.   *Unstressable word syndrome* (Hayes, 1995: §5.3, 402)

         Extrametricality is blocked if
                  (i) the language has a strong ban on degenerate feet,
         and   (ii) the application of extrametricality creates a subminimal domain.

Note that these exception clauses handle interactions between parameters in purely stipulative ways: in (8b), in particular, provision (i) establishes an *ad hoc* link between the suspension of extrametricality and a strong setting of the foot degeneracy parameter, whilst the applicability of provision (ii) cannot be evaluated without knowing whether the language affords alternative repair strategies such as lengthening. In OT, in contrast, no special stipulation is needed to predict extrametricality blocking (see e.g. Bermúdez-Otero, 1998: 184-5); indeed, the various strategies that languages deploy against the unstressable word syndrome (Hayes, 1995: §5.3) fall out by factorial typology from the interaction of maximally simple constraints enforcing foot binarity, extrametricality, moraic faithfulness, and culminativity.

In syntax, it has long been obvious that hedges cannot be avoided in theories that aim to work exclusively with inviolable rules and constraints. Speas (1997: 184) provides a list of twelve such hedges. In some of Speas's examples, the need for a hedge arises from a particular set of assumptions about the nature of syntactic representations, rather than from the mere use of inviolable rules or parameters. Nonetheless, there are many syntactic phenomena for which it is difficult to see how inviolable rules could be formulated in any framework without recourse to hedges. A case in point is the crosslinguistic variation in the use of zero pronouns in subject positions, often analysed in terms of a pro-drop parameter.[18] In simple terms we can formulate the parameter with its full complement of (language-specific) hedges as follows:

(9)   **The pro-drop 'parameter'**

Use a zero pronoun in subject function
• but only if there is inflection to support the zero element,[19]
• but only if pro is a topic,
• but even then only if the antecedent is in a structurally acceptable position.

The analyses of this phenomenon provided by Rizzi (1986) and James Huang (1984, 1989) clearly illustrate the difficulties incurred by theoretical frameworks that only countenance inviolable principles and parameters; for criticism of these treatments, see e.g. Yan Huang (1992, 1995a, 1995b). Pro-drop phenomena have since been subject to a number of successful optimality-theoretic analyses: e.g. Samek-Lodovici (1996), Cole (2000), Speas (2001).[20]

4.3. Emergence-of-the-unmarked effects

Accounting for the emergence of the unmarked (McCarthy & Prince, 1994) is perhaps the clearest advantage of OT's appeal to universal violable markedness constraints. The term 'the emergence of the unmarked' refers to the prediction that, in contexts where lexical specifications have no effect, only relatively unmarked structures occur. This prediction follows straightforwardly from the principle of minimal violation: in the absence of faithfulness requirements, the highest-ranked relevant markedness constraint must necessarily be satisfied.

Research has discovered numerous instances of the emergence of the unmarked in phonology. Segment epenthesis provides a good example (see e.g. Kager, 1999: §3.5): since an epenthetic segment has no input correspondent, faithfulness will have no effect on its featural content, which will therefore be determined by markedness constraints.[21] In the case of vowels, for example, markedness constraints such as those in (10a) can be postulated on the basis of evidence from implicational universals governing vowel inventories. As predicted by the emergence of the unmarked, typical epenthetic vowels such as [i, e, ɨ, ə] obey these independently established constraints, whereas vowels that violate them tend not to occur as epenthetic segments:[22]

(10)   **The emergence of the unmarked in vowel epenthesis**

a. *Markedness constraints on vowels*	b. *Dispreferred epenthetic vowels*
*[+round, –back]	y, ø
*[+low, –back]	æ
*[+round]	u
*[+low]	a, ɑ

Interestingly, emergence-of-the-unmarked effects are also obvious in syntax. Bresnan's (2001b) analysis of pronominal systems provides a good example. In general, three types of properties can be associated with pronouns:

(11) **Properties associated with pronominal elements**

a.　　　Semantically, they are characterized by having variable reference.

b.　　　In terms of information structure, they may show a specialization for reference to topical elements.

c.　　　Morphosyntactically, they can bear marking for agreement features.

As regards their form, pronominal elements range from zero to full-pronoun status, with three intermediate possibilities:

[INSERT (12) HERE]

A pronominal element can thus be thought of as a pairing between the features in (11) and the forms in (12). However, not all possible pairings occur in language and, among those that are possible, some are typologically more marked than others. These restrictions can be captured by means of two correspondence constraints and three structural markedness constraints (Bresnan, 2001b: 119-23):

(13) **Constraints on pronominal elements**

a.　　　*Correspondence constraints*
　　　　(i)　REDUCED⇔TOPIC
　　　　　　Pronominals are reduced if and only if they are specialized for topic anaphoricity.
　　　　(ii)　OVERT⇔AGREEMENt
　　　　　　Pronominals are inherently specified for person/number/gender if and only if they are overt.

b.　　　*Structural markedness*
　　　　(iii)　FULL
　　　　　　Pronouns are full.

　　　　(iv)　ICONICITY
　　　　　　No zero or bound pronouns.

　　　　(v)　ALLOTAXY
　　　　　　No weak or clitic pronouns.

In those languages where the structural markedness constraints in (13b) rank high, full pronouns surface even in topic function: see e.g. Bresnan (2000b) on the pronominal systems of pidgins. However, if some structural markedness constraint is demoted below an antagonistic correspondence constraint, then the optimal output in topic functions will be a reduced form: e.g. Chicheŵa. Even in such systems, however, the structurally unmarked full form can emerge to fill a paradigmatic gap. In Chicheŵa, for example, prepositions combine with reduced forms of pronouns when the latter express topic; the free pronoun is reserved for emphasis and contrast.

(14) a. ndí   íwo   b. nǎwo
         with it   with.it[TOP]

However, some prepositions lack the bound form; in such cases, the unmarked free pronoun surfaces, even if it is a topic.

(15) a. kwá íyo   b. *kwǎyo
         to   him

Emergence of the unmarked effects in pronominal systems have also been observed in the evolution of the Romance languages from Latin (Vincent, 2000). For the emergence of the unmarked in other syntactic phenomena related to word order, see e.g. Costa (2001) and Müller (2001).

5. Empirical difficulties for OT

So far our discussion has shown that the types of facts which OT handles best occur both in phonology and in syntax. We now turn to phenomena that raise difficulties for the theory: ineffability (§5.1), optionality (§5.2), and allegedly arbitrary patterns arising from diachronic processes of grammatical restructuring (§5.3). As will become apparent in the course of this and further sections, these problems are not all equally serious, and we do not feel, in any case, that they warrant abandoning OT in favour of any of the currently available alternatives. Nonetheless, it is not our intention here to provide solutions, but rather to demonstrate that these problems are common to phonology and syntax, and do not therefore provide an argument for confining the application of OT to the former. As in §4, we focus on empirical difficulties arising from OT's reliance on universal violable markedness constraints, rather than from other features of the theory such as parallelism.

5.1. Ineffability

The terms 'absolute ill-formedness' (Prince & Smolensky, 1993: 48; McCarthy, 2002: §4.1.2), 'absolute ungrammaticality' (Kager, 1999: §9.3), and 'ineffability' (Pesetsky, 1997) designate a situation where there is no grammatical output for a given legitimate input. Ineffability is problematic for OT because, for any licit input, GEN necessarily emits a nonempty candidate set, and, within this candidate set, there always is at least one candidate that best satisfies the constraint hierarchy. According to the basic postulates of OT, such a candidate must constitute the grammatical output.

Pesetsky (1997, 1998) relies upon syntactic ineffability to argue that, whilst OT can deal with the phonological realization of syntactic structures ('pronunciation'), core syntactic phenomena such as movement do not operate optimality-theoretically but rather in a 'clash-and-crash' fashion; see §3 above. In Pesetsky (1997: §2), he notes for example that, in English, the presence of an overt subject renders embedded infinitival *wh*-questions absolutely ungrammatical, as all possible realizations incur some fatal constraint violation:

(16) **Ineffable embedded infinitival *wh*-questions with overt subjects**

a.   *Mary wonders [Bill to buy which book]   (*wh*-phrase *in situ*)
b.   *Mary wonders [which book for Bill to buy]   (doubly filled Comp)
c.   *Mary wonders [which book Bill to buy]   (*Bill* not in Case position)

This particular case, however, does not seem obviously intractable if we assume that syntactic mappings may be unfaithful by, for example, inserting tense or modality features absent from the input (see e.g. McCarthy, 2002: 198-200, and references therein). Under this assumption, the ungrammatical structures in (16) simply lose to unfaithful competitors such as *Mary wonders which book Bill will buy*. If so, (16) only shows that, in English, a constraint such as DEP-tense ('Do not insert tense features') cannot be superordinate; the feasibility of the analysis depends on whether or not this ranking can be maintained across constructions.

Faithfulness-based solutions to ineffability effects have been advocated in Legendre *et al.* (1995) and Legendre, Smolensky & Wilson (1998). These authors deal with the absolute ungrammaticality of multiple *wh*-questions in some varieties of Italian by arguing that, in the winning output candidate, one of the *wh*-features of the input is left unparsed and the corresponding *wh*-argument replaced by an indefinite pronoun. Of course, this type of faithfulness-based solution to ineffability effects forces one to reject the 'Katz-Postal Hypothesis for GEN' (McCarthy, 2002: 198; the term refers to proposals made in Katz & Postal, 1964), according to which all the members of a candidate set must share the same interpretation —i.e., in GB or Minimalist terms, be associated with the same LF representation.[23] Abandoning the Katz-Postal hypothesis for GEN is not without perils. Kuhn (2003) argues that syntactic generation becomes undecidable if candidates are allowed to vary arbitrarily far from input semantics. Kuhn proposes an LFG-OT solution to this problem: he stipulates that the f-structure of a candidate must preserve the semantic content of the input, but allows for unfaithfulness in the form of divergence between the global f-structure of a candidate, on the one hand, and the candidate's c-structure and lexical entries, on the other.

Adhering to the Katz-Postal Hypothesis for GEN, Ackema & Neeleman (2000) deal with absolute syntactic ungrammaticality by means of the notion of 'null parse'. In this view, an expression is ineffable when its optimal output realization is ∅ (Prince & Smolensky, 1993: 48). However, using the null parse in syntax requires a few delicate stipulations. First, Ackema & Neeleman claim that the null parse is included in the candidate set for every syntactic input because, having no semantics, it cannot have a deviating interpretation. Naturally, the null parse violates faithfulness constraints profusely, but, as long as these are ranked low, it can win. Ackema & Neeleman (2000) furthermore assume that faithfulness constraints can only refer to features: if a candidate fails to include a lexical element present in the input, this will cause the candidate to have the wrong semantics, rather than to incur a faithfulness violation. Nonetheless, whether or not these stipulations are regarded as suspect, the null parse is unlikely to solve the issue of absolute syntactic ill-formedness, as it has proved inadequate even in phonological applications (see below).

Whatever the correct approach to syntactic ineffability proves to be, the existence of absolute ungrammaticality in phonology refutes Pesetsky's claim that ineffability distinguishes core syntax from the realm of 'pronunciation'. Phonological ineffability is evident in cases of prosodically or phonotactically conditioned morphological defectiveness, in which a legal combination of morphemes is left unrealized in order to avoid a marked phonological configuration. In some Turkish dialects, for example, morphologically derived forms are subject to a disyllabic minimality constraint, as a result of which combinations of a /(C)V-/ root with a /-C/ suffix are absolutely ill-formed:

(17)   **Ineffable suffixed monosyllables in Turkish (Inkelas & Orgun, 1995: §4)**

a.   kafa-m   'head-1SG.POSS'
b.   fa-dan   'F[the musical note]-ABL'
c.   *fa-m   'F[the musical note]-1SG.POSS'

The question arises as to why the missing forms are not rescued by repair strategies such as epenthesis: e.g. /do-m/ → *[dom], *[do.**u**m], *[do.**ju**m] 'C[the musical note]-1SG.POSS'. As Orgun & Sprouse (1999) show, one cannot simply claim that the null parse is less disharmonic than an epenthesized realization: epenthesis proves preferable to the null parse as a means to avoid hiatus and illicit coda clusters.

(18)   a.   /araba-a/   →   [a.ra.ba.**ja**]   'car-DAT'
        b.   /it-m/   →   [i.t**i**m]   'dog-1SG.POSS'

One might be tempted to speculate that the null parse and epenthesis are optimal in different cophonologies (see e.g. Orgun, 1996) associated with different morphological constructions; (18b), however, shows that the same suffix that causes ineffability with /(C)V-/ roots triggers epenthesis after consonant-final bases. In view of these difficulties, Orgun & Sprouse (1999) propose supplementing OT with a theory of CONTROL, but it is unclear whether even this solution achieves full generality (see e.g. Törkenczy, 2000).

5.2. Optionality

The concept of optionality is straightforward enough: optionality occurs whenever a computation assigns two or more distinct but equally well-formed outputs to a given input. Identifying genuine empirical instances of optionality is, however, a tricky and theory-laden affair, for linguistic variants are seldom (if ever) equivalent in all respects. Accordingly, asserting that the choice between *a* and *b* is, say, *syntactically* optional is tantamount to claiming (i) that *a* and *b* have strictly identical representations in the input to the syntactic computation, and (ii) that the choice between *a* and *b* is effected not by the syntax but by some other linguistic or nonlinguistic cognitive system. These are not trivial claims. (For similar remarks, see McCarthy, 2002: 200.)

Insofar as it exists, true optionality is problematic for OT. Assuming the relative ranking of all conflicting constraints to be fixed, two output candidates can be equally harmonic if, and only if, they incur strictly identical sets of constraint violations. The likelihood that this will occur is inversely proportional to the richness of CON, but research trends over the past decade suggest that considerable numbers of constraints need to be postulated if OT is to achieve descriptive adequacy. Thus, global ties among competing candidates can be expected to be vanishingly rare. In this section we shall not attempt to ascertain to what extent true optionality exists either in phonology or in syntax. We will merely note that similar strategies seem to be appropriate in both fields to deal with what McCarthy (2002: 200) calls 'observational optionality', which may or may not be 'true'.

One way of modelling true optionality is to relax the assumption that all crucial rankings are fixed. Thus, Boersma (1997, 1998a) and Boersma & Hayes (2001) advocate the notion of *stochastic evaluation*: individual constraints are given a probability distribution along a linear scale of constraint strictness; when the probability distributions of two constraints c_i and c_j overlap, both $c_i » c_j$ and $c_j » c_i$ rankings will occur – with different probabilities – in different instances of evaluation. Thus, stochastic evaluation captures some of the essential properties of Labovian variable rules. Boersma & Hayes (2001) apply these

ideas to the analysis of optional phonological processes of metathesis and reduplication in Ilokano. Interestingly, Bresnan, Dingare & Manning (2001), cited by Bresnan & Aissen (2002), argue that stochastic evaluation should also be used to account for syntactic variation. They observe that, in Lummi, the subject cannot be lower in the person hierarchy than the nonsubject argument; such marked configurations are avoided by passivization, which enables the argument that is higher in the person hierarchy to be promoted to subject position. In English, passivization is optional: instances of marked person/argument alignment are permitted, but statistical corpus analysis reveals that the incidence of passives increases in contexts where the marked configuration would otherwise occur (Aissen 1999). Accordingly, the same ranking holds categorically in Lummi, yielding an obligatory pattern, and stochastically in English, yielding a truly optional pattern.[24]

Alternatively, one can deal with observational optionality by denying that it is true, asserting instead that it merely reflects a choice between two different inputs. That certain instances of phonological optionality must be so analysed is uncontroversial: in British English, for example, the choice between /skɒn/ and /skəʊn/ 'scone' and /ʃrəʊzbərɪ/ and /ʃruːzbərɪ/ 'Shrewsbury' is undeniably made in the lexicon, not in the phonology. In a similar vein, Baković & Keer (2001) assert that, in English, the presence or absence of the complementizer *that* in *I think (that) the coat won't fit him* is determined in the input to the syntactic computation, and reflected in the output by simple faithfulness. Stipulative though this is, it is worth noting that this is essentially the strategy assumed in Minimalist syntax, where different syntactic options reflect different Numerations (see Burton-Roberts & Poole, this volume). More generally, true optionality will be rare in any computational system making extensive use of defaults; this fact should be borne in mind when assessing Newmeyer's (2000) arguments against optimality-theoretic approaches to syntactic optionality.

5.3. Grammatical restructuring and 'crazy rules'

In the course of language change, extragrammatical performance phenomena often give rise to new phonological and syntactic patterns through processes of phonologization and grammaticalization.[25] These new grammatical patterns typically comply with universal markedness generalizations and are, in this sense, 'natural'. After their creation, however, the innovated patterns continue to evolve and may be affected by diachronic processes of restructuring. The critics of OT have seized upon this fact, claiming that the outcome of grammatical restructuring is often 'unnatural' and cannot be described in terms of the interaction of markedness constraints (see among others Hale & Reiss, 2000; McMahon, 2000). This, they claim, proves that grammars do not consist of hierarchies of markedness constraints and that markedness generalizations are mere epiphenomena of recurrent processes of phonologization and grammaticalization (see further §6, specially §6.1).

In early generative phonology, this idea was forcefully articulated by Bach & Harms (1972), who argued that innovation creates 'plausible' rules, which may in the course of 'grammar simplification' (i.e. restructuring) become 'crazy'. A form of diachronic restructuring alleged to produce arbitrary phonological patterns is rule telescoping, whereby a series of processes arising from successive sound changes merges into an innovative single rule that is not phonetically motivated. In Tswana, for example, the collapse of a sequence of separate innovations has given rise to obstruent devoicing after nasals: e.g. /N+b/ → [mp] (Hyman, 2001; see further Myers, 2002); this process is unnatural, in that it directly contradicts the markedness constraint *NC_{[-voi]} (see §4.1).

By far the largest share of attention, however, has been devoted to rule inversion (Vennemann, 1972).[26] Rule inversion is triggered by the reanalysis of alternations: given an

input element /α/ with two alternant output realizations [α] and [β] respectively occurring in the environments [__]ₐ and [__]_b, inversion involves the replacement of the old mapping /α/ → [β] in environment [__]_b by a new mapping /β/ → [α] in environment [__]ₐ. In nonrhotic dialects of English, the postlexical process of *r*-intrusion inserting [r] into empty onsets after (nonhigh) vowels arose from the inversion of a rule deleting /r/ in the coda. Since [r] is a relatively marked segment, *r*-intrusion is claimed to be unnatural and to disprove the claim that the quality of epenthetic segments is determined by emergence-of-the-unmarked effects (see §4.3). This – the critics of OT assert – shows that grammatical restructuring is not constrained by markedness; we will however refute this argument in §6.5.

In syntax too, diachronic restructuring gives rise to grammars containing apparently arbitrary patterns. Lightfoot (1999) discusses several instances, which he colourfully describes as 'Rube Goldberg features of grammars'. One of the examples he provides is the rise of ergative-absolutive case systems. From the viewpoint of grammatical relations, ergative morphology may appear puzzling: first, it involves a departure from the usual pattern of correspondence between syntactic function and morphological case (subject→nominative, object→accusative); secondly, it normally coexists with nominative-accusative marking in so-called 'split ergative systems'. As S. Anderson (1977, 1992: §13.3.1) has shown, however, this strange phenomenon can be understood as the product of historical reanalysis. Thus, one of the scenarios leading to the rise of ergativity involves the reanalysis of passive constructions as active transitive structures: as the oblique agent phrase of the passive becomes the subject of the new active construction, its agentive marking is reinterpreted as ergative case:

(19)   **The reanalysis of passive syntax as ergative morphology**

	V	NP	PP/NP
a. *syntactic category*	V	NP	PP/NP
semantic rôle	—	patient	agent
syntactic function	—	subject	adjunct
marking	passive	(nominative)	agentive preposition/marker
	↓	↓	↓
b. *syntactic category*	V	NP	NP
semantic rôle	—	patient	agent
syntactic function	—	object	subject
marking	active	(absolutive)	ergative marker

A good example of this development can be found in Polynesian (Chung, 1978). A comparison of conservative Maori (20a) with innovative Tongan (20b) shows how *-Cia*, originally a passive verbal suffix, ceased to denote passive voice; concomitantly, the agentive preposition *e* became an ergative case marker.

(20)   **The rise of ergativity in Polynesian**

a.   *Maori (= proto-Polynesian)*

  active   Ka   whana te hōiho i   a   Hōne
          TENSE kick   the horse ACC the.PROPER John
          'The horse kicked John'

  passive   Ka   whana-a   a   Hōne **e** te hōiho
          TENSE kick-*Cia*(PASS) the.PROPER John **by** the horse
          'John was kicked by the horse'

b. *Tongan*

    active   Ne'e fangu-na   au 'e he nanamu 'o e kakalá

              PAST awaken-*Cia*(ACT) me ERG the smell   of the flower

              'The smell of the flower awakened me'

    no syntactically derived passive

According to S. Anderson (1992: 354), this reanalysis was motivated by the fact that, in Polynesian, the passive construction has a particularly wide distribution and is obligatory in certain contexts; in consequence, the syntactically derived status of passives became opaque in languages like Tongan.

    Although these examples are striking, the unnaturalness of the grammatical patterns created by diachronic restructuring has nonetheless been exaggerated, both in phonology and in syntax. As we noted above, later sections (§6.5) will show that phonological rule inversion is constrained by markedness. Similarly, ergative case systems can be motivated synchronically in terms of discourse organization (DeLancey, 1981; Du Bois, 1987); Hopper & Traugott (1993: 160-3) provide a brief summary of the argument.

6. Grounding: refuting the charge of 'substance abuse'
6.1. Does grounding violate Ockham's Razor?

In §4 and §5 we have examined the implications of OT's reliance upon universal violable markedness constraints. The results are unequivocal: in no case has a prediction of OT been confirmed in phonology but refuted in syntax, or vice versa; the advantages and problems of universal violable constraints clearly cut across the phonology/syntax divide. Syntax, therefore, is as likely as phonology to be governed by markedness. In §3, however, we argued that markedness constraints are *a* (if not *the*) locus of grounding. If so, one must expect grounding to be as much an attribute of syntax as it demonstrably is of phonology.

    The concept of grounding is nonetheless surrounded by serious problems. Notably, OT has been argued to incur massive redundancy (if not downright incoherence) by according cognitive reality to linguistic markedness generalizations whilst acknowledging them to be grounded in nonlinguistic substance. Hale & Reiss (2000), Hale (2000), and Reiss (2000) – henceforth Hale & Reiss – have forcefully pressed this charge under the memorable title of 'substance abuse'. Similar views are expressed by critics of OT both in the functionalist and in the formalist camps: e.g. Haspelmath (1999), McMahon (2000), Newmeyer (2000; 2002a, b). In essence, they all assert that, by representing markedness generalizations in the grammar as constraints, OT incurs gross violations of Ockham's Razor.

    This accusation is supported with several arguments. First, it is claimed that, *qua* cognitive entities, markedness constraints are causally inert: grammatical constraints are not needed to enforce universal markedness generalizations because these, it is argued, emerge diachronically from recurrent processes of phonologization and grammaticalization driven by performance factors (see §5.3). Consider, for example, the following implicational universal:

(21)   In every language, the onset supports no fewer place-of-articulation oppositions than the coda.

This statement holds true because, in general, perceptual cues to consonantal place are less reliable for codas than for onsets (see §2); in consequence, coda consonants are subject to higher misperception rates, and so diachronic processes of phonologization giving rise to

phonological rules of place neutralization affect codas much more frequently than onsets. In this causal scheme, a constraint such as CODACOND-Place, as stated in (1), plays no rôle.

This argument is reinforced by the claim that gaps in factorial typology are also diachronically epiphenomenal. Myers (2002), for example, suggests that, although no attested language uses vowel epenthesis to break up marked NC$_{[-voi]}$ sequences (see §4.1), it would be wrong to tinker with the composition of CON in an attempt to predict this fact. As a phonological process, Myers argues, vowel epenthesis in NC$_{[-voi]}$ clusters is *formally possible*, but there simply is no plausible sequence of phonetically motivated changes that can historically give rise to it. The critics of OT (among whom Myers is interestingly not included) would use this result to underpin their claim that attempts to bring typological principles into the sphere of cognition are fundamentally wrong-headed (Newmeyer 1998).

Although Hale & Reiss describe OT as *conceptually* incoherent, however, their own approach to the nature and origins of markedness embodies a set of *empirical* claims; one can logically conceive of alternative scenarios, where cognitive representations of markedness (in the form of optimality-theoretic constraints) are not causally inert. First, we may hypothesize without logical contradiction that performance-driven change has the potential to create both 'natural' and impossible linguistic patterns; an impossible pattern will however not be acquired if it cannot be represented by any ranking of the universal constraint set. Similarly, since historical changes often proceed in opposing directions, it is perfectly conceivable that, under particular circumstances, a sequence of externally driven changes could give rise to a grammar that generates marked structures but prohibits their unmarked counterparts, in violation of some implicational universal (see again the discussion of rule telescoping in §5.3). In OT, however, this development would be impossible because, if a structure satisfies all markedness constraints in CON,[27] then it constitutes a possible output whether or not it occurs in the learner's ambient language; conversely, to acquire a marked element the learner must demote some markedness constraint on the basis of positive evidence in the primary linguistic data.[28] It is thus logically possible that, whilst diachronic processes bear the main burden of filtering marked structures out of language, the precise form of crosslinguistic markedness generalizations is determined by the interaction of such processes with grammatical constraints, which, through their rôle in language acquisition, prove to be causally efficient. In §6.2 to §6.5 we flesh out this hypothesis with empirical arguments.

Why, then, do Hale & Reiss refuse to countenance such scenarios? As it turns out, the claim that markedness constraints are causally inert depends on a more basic contention: namely, that linguistic patterns governed by markedness do not raise Plato's Problem in a serious form or, in other words, that markedness relationships are not subject to poverty of the stimulus. If this is true, then UG need not supply markedness constraints; from the viewpoint of learnability, they are superfluous. Hale (2000: 250ff.) makes his position clear by means of a thought experiment. Hale invites us to imagine the linguistic fate of a genetically mutant human creature that he labels *homo collitumens*. Cognitively, *homo collitumens* does not have a constraint against voiced obstruents (VOICEDOBSTRUENTPROHIBITION, or 'VOP'; see e.g. Kager, 1999: 40) or against voice contrasts in the coda (CODACOND-voice; cf. CODACOND-Place in (1) above). Physiologically, the mutations hypothesized by Hale remove all phonetic impediments to obstruent voicing. Thus, for *homo collitumens* the marked status of voiced obstruents is neither cognitively represented nor phonetically grounded. Nonetheless, Hale (2000) asserts that *homo collitumens* infants would have no difficulty in acquiring languages such as Hawaiian (where voiced obstruents are altogether prohibited) or German (where obstruents are devoiced in the coda); the mutant children would learn these processes effortlessly *by induction over the input data* (Hale, 2000: 252). On this point, a formalist such as Hale and a functionalist such as Haspelmath find themselves in agreement: Haspelmath (1999: §5) asserts that markedness-

governed patterns become established in grammars by a process of 'cognitive entrenchment' (Langacker, 1987), which we take to designate an essentially empiricist inductive mechanism.

Hale & Reiss's arguments are cogent and, if tenable, devastating for OT; but, as we have carefully pointed out, they are vulnerable to empirical disconfirmation. Thus, a theory of grammar that admits grounded markedness constraints will turn out not to violate Ockham's Razor if one can show the following:

(i) that the markedness generalizations attested by linguistic typology and embodied in OT constraints differ in specifiable ways from the linguistic patterns predicted to emerge from mere induction over linguistic data sifted through the Arena of Use;[29]

(ii) relatedly, that infants rely on markedness constraints to transcend the limitations of induction over primary linguistic data (where 'primary linguistic data' designates exposure to a series of utterances in the ambient language).

Both propositions, we submit, are empirically true; we show this in §6.2 to §6.5. To simplify matters, our discussion will focus on phonology, where the issues are clearest and arguments against markedness constraints have been formulated most sharply (see §2).

6.2. Difficulty landscapes vs markedness constraints[30]

Consider the marked status of voiced obstruents, to which we alluded in §6.1. This, as is well known, is grounded on the fact that it is difficult to maintain vocal fold vibration if egressive pulmonic airflow is restricted supraglottally (Ohala, 1983; Westbury & Keating, 1986). Vocal fold vibration becomes harder to maintain (i) the longer the supraglottal constriction is held and (ii) the farther back in the oral cavity the constriction is located. The voiced plosives [b, d, g, bː, dː, gː] can accordingly be arranged in the following performance difficulty landscape:

[INSERT (22) HERE]

In (22), performance difficulty increases monotonically from top to bottom and from left to right: within any column, the segment in the top cell is easier than the segment in the bottom cell; within any row, any segment is easier than those occupying cells further to the right. Thus, [d] is easier than [dː] and [g] and, by transitivity, than [gː]. The concept of relative performance difficulty can be operationalized, we assume, in terms of statistical rates of successful production and recognition in an appropriate experimental setting.

According to Hale & Reiss, a linguistic structure is marked if it is often lost through historical change, and the probability that a structure will become subject to historical loss depends in turn on its relative performance difficulty. This argument leads to the following prediction: in any given set of segments, those above any arbitrarily defined threshold of relative performance difficulty t will be more marked than those below t; in other words, linguistic markedness hierarchies are isomorphic with performance difficulty landscapes.

Now consider again the difficulty landscape in (22). If we establish a performance difficulty threshold t immediately below [d], there are two possible partitions of the segment set: if [d] is easier than [bː], we obtain (23b); otherwise (23a) results.

[INSERT (23) HERE]

Empirically, the choice between (23a) and (23b) depends on the respective contributions of closure duration and closure location to the overall performance difficulty of a voiced plosive: (23b) will be correct if – and only if – the contribution of closure duration exceeds that of closure location by a sufficiently large margin; otherwise, (23a) will be true. By Hale & Reiss's arguments, however, landscape (23a) gives rise to an impossible linguistic markedness hierarchy (following established usage, ≻ means 'more harmonic than'):

(24)   { b, bː, d } ≻ { dː, g, gː}

Observe that, in (24), the segment sets defined by the markedness threshold do not constitute natural classes: e.g. no language has a phonological rule devoicing /dː, g, gː/ but not /b, bː, d/. Therefore, if one asserts that linguistic markedness hierarchies are isomorphic with performance difficulty landscapes, one runs the risk of predicting impossible markedness generalizations.

Our example illustrates the risks of Hale & Reiss's position but falls short of delivering an actual empirical counterargument, for we have not shown whether (23a) or (23b) is in fact correct. Other studies, however, have clearly proved that linguistic markedness hierarchies do *not* match performance difficulty landscapes. Notably, using a software aerodynamic model of the vocal tract (Keating, 1984), Hayes (1999a: §6.2) derived the impossible markedness hierarchy given in (25). The anomalous portion of the hierarchy is highlighted in bold typeface; as above, ≻ means 'more harmonic than', whilst ‿ denotes strict linear precedence.

(25)   { [+nas]‿b, [+nas]‿d, [+nas]‿g } ≻ **[+son, −nas]‿b ≻ [+son, −nas]‿d ≻ #‿b ≻ #‿d ≻ [+son, −nas]‿g** ≻ #‿g ≻ [−son]‿b ≻ [−son]‿d ≻ [−son]‿g

In OT, monstrosities such as (24) and (25) are straightforwardly avoided:

First, constraints possess the attribute of self-containment: i.e. they are formulated in terms of specifically linguistic categories (see §2). Phonological markedness constraints accordingly refer to distinctive features (which pick out natural classes of segments) rather than to phonetically defined segment lists.

Secondly, markedness constraints consist of relatively simple statements and cannot therefore reproduce the full complexity of performance effects. Consider, for example, the phonological constraints that require consonants to be coronal when adjacent to a front vowel. These constraints are grounded upon the fact that, in such contexts, labials and dorsals are often misperceived as coronals (see e.g. Ohala, 1989: 182-5). Cole & Iskarous (2001), however, show that misperception rates are highest when a consonant is flanked by front vowels on both sides. Despite this phonetic fact, the phonological constraints that drive coronalization never require a 'two-sided environment'; rather, the presence of a following

(or sometimes preceding) front vowel suffices to trigger coronalization. In other words, the relevant markedness constraints refer to relatively simple one-sided environments such as __V$_{[-bk]}$, rather than to more complex two-sided contexts such as V$_{[-bk]}$__V$_{[-bk]}$ (see further §7.2).

Thirdly, constraints do not interact additively, but are subject to language-specific relationships of strict dominance. For this reason, grammars do not reflect the trade-offs that characterize the interaction between different sources of difficulty in performance (Hayes, 1999a: §6.2).

In sum, there are clear mismatches between linguistic markedness generalizations and performance difficulty landscapes. In standard OT this fact follows from the assumption that constraint sets are self-contained, that constraint statements are relatively simple, and that constraints interact through language-specific dominance hierarchies.

6.3. Performance environments vs structural positions

We have argued that linguistic markedness cannot be reduced to performance difficulty because the former refers to I-language categories whilst the latter is a property of E-language entities; this is reflected in the self-contained character of optimality-theoretic constraints. Markedness constraints referring to structural relationships such as syllabic constituency provide further examples of the same idea.

Consider again CODACOND-Place. This constraint prevents rhyme positions from licensing a feature-geometric C-Place node. There can be no doubt that such a constraint evaluates phonological representations, rather than phonetic events. For a consonant, affiliation to the rhyme is a purely phonological property: in phonetics there are no syllables, onsets, or rhymes, but only articulatory gestures, acoustic spectra, auditory percepts, and the like. In fact, not only are phonetic events and phonological representations ontologically distinct, but in the case of syllables there is also no language-general algorithm determining the correspondence between the former and the latter: as Ladefoged & Maddieson (1996: 281-2) point out, there is no phonetic parameter that can be used to define syllabicity (and the same is probably true of segmenthood and of foot boundaries). In consequence, a constraint such as CODACOND-Place cannot be reduced to – or deduced from – any collection of statements about phonetics because CODACOND-Place refers to entities that neither are phonetic nor can be defined phonetically.

In recent years, however, many phonologists working in OT have argued that markedness constraints refer directly to phonetic substance (e.g. Kirchner, 1997; Steriade, 1997, 2000; Zhang, 2000).[31] The line of reasoning leading to this position can be illustrated with an example from Klamath discussed by Steriade (see Gerfen, 2001: 186). Steriade notes that, in Klamath, the neutralization of aspiration and ejection contrasts among obstruents does not depend on syllable structure, but reflects purely linear constraints: laryngeal oppositions are neutralized before obstruents and word finally. In /p^{h}et'-wa/ → [p^{h}et'.wa] 'floats in water', the [t'], though rhymal, supports distinctive ejection because it precedes a sonorant. According to Steriade, what renders presonorant environments special is the availability of appropriate perceptual cues for aspiration and ejection, particularly burst phenomena and VOT (Voice Onset Time); she therefore argues that constraints regulating the distribution of aspiration and ejection should directly refer to these properties of the phonetic environment.

These conclusions are however not warranted by the evidence. In particular, it is far from clear why the constraint exemplified by Klamath cannot be stated in purely phonological (albeit linear) terms, i.e. as requiring an immediately following sonorant – where [+son] is a phonological specification. As we have shown, moreover, constraints such as CODACOND-Place cannot be reformulated in terms of properties of the phonetic environment because phonological entities such as the coda lack a language-general phonetic

definition. Gerfen (2001) has recently provided new empirical evidence that reference to syllable structure in such constraints is indispensable. In certain dialects of Spanish, for example, obstruents undergo debuccalization and aspiration to [h] precisely in the coda; no set of phonetic variables adequately characterizes the environment of these processes. Thus, one cannot say that place and voice neutralization are triggered by the absence of essential perceptual cues provided by a "right-hand modal sonorant context" (Steriade, 1997: 94): in Spanish, /s/ is rhymal before /l/, and therefore subject to aspiration even though /l/ is a sonorant. In response, one could argue that in homorganic clusters such as /sl/ segmental transitions provide relatively impoverished phonetic cues; but, in that case, one would not be able to explain why /tl/ escapes neutralization precisely in those aspirating dialects in which the cluster is tautosyllabified. A similar state of affairs obtains in Catalan, where the laryngeal features of a plosive immediately followed by a liquid are contrastive if the cluster is tautosyllabic, but predictable if the cluster is heterosyllabic: e.g. [sem.pɾə] 'always' vs [sem.bɾə] 'sow.3SG.PRES'; cf. [sub.ru.ti.nə] 'subroutine', *[-p.r-] (Wetzels & Mascaró, 2001: note 21).

Additionally, direct reference to phonetic substance raises insurmountable obstacles to the analysis of categorical phonological processes. If, for example, constraints enforcing neutralization referred directly to the perceptual cues available in the phonetic environment, then neutralization would often fail in hyperarticulated speech, which is by definition particularly rich in perceptual cues (Lindblom, 1990); but, in fact, categorical neutralization processes apply regardless of speaking style (Myers, 2000: 266; see also Hayes, 1999a: §6.1).

In sum, markedness generalizations refer to I-language categories. Such categories are ontologically distinct from E-language entities, and there often is no consistent correspondence between the two sets. Despite being largely nonarbitrary, therefore, optimality-theoretic constraints cannot be logically reduced to statements about entities in the realm of performance.

6.4. Markedness constrains performance-driven change

In §6.2 and §6.3 we have showed that linguistic markedness generalizations differ in crucial ways from the patterns that would emerge by mere induction from linguistic data sifted through the Arena of Use. Standard OT captures these differences by postulating self-contained constraint sets, by formulating constraints in relatively simple terms, and by defining strict dominance as the basic mode of constraint interaction. This proves that optimality-theoretic constraints cannot be reduced to statements of performance difficulty. Nonetheless, one may argue that, even without cognitive representations of markedness, a linguistic theory relying heavily upon induction can predict the observed mismatches between linguistic markedness relationships and performance difficulty landscapes if (i) it provides an appropriately restrictive representational format for linguistic rules and (ii) it incorporates an evaluation measure[32] with a bias for formally simple rules. In such a theory, markedness asymmetries are still assumed to arise by induction over primary data subject to performance distortion, but the theory of representations and the simplicity measure take care of the mismatches between linguistic generalizations and performance phenomena. In essence, the rule-based nonlinear phonological theories of the 1980s (minus markedness theory) fit this description. It remains to be shown, therefore, that knowledge of markedness relationships is essential to the acquisition process; in §6.4 and §6.5 we turn to this task.[33]

Following a line of reasoning sketched in §6.1, we shall first argue that, if language learners did not enjoy access to knowledge of markedness relationships, performance-driven change could give rise not only to natural linguistic patterns but also to impossible linguistic systems. An argument to this effect was first formulated by Jakobson (1929) and has recently

been echoed by Kiparsky (1995) and Bermúdez-Otero & Hogg (2003). Jakobson observed that, according to the Neogrammarians, sound change is caused by phonetic factors such as ease of articulation and operates 'blindly', i.e. without regard for its impact upon the linguistic system. Against this view, Jakobson objected that sound change cannot be totally blind, for otherwise it would violate the limits imposed by implicational universals. This possibility can be illustrated with an example from Kiparsky (1995: 641). Historically, voiceless stops can be lost through processes of affrication and spirantization driven by phonetic phenomena. Insofar as these phonetic phenomena can occur in a language regardless of the size and composition of its phoneme inventory, phonetically motivated sound change could give rise to a grammar forbidding voiceless stops altogether. Such a development, however, is impossible, since voiceless stops are the least marked obstruents and, as such, their presence is required by implicational universals. Markedness principles are therefore capable of blocking phonetically motivated sound changes when the latter would have unacceptable global effects.

The Jakobson-Kiparsky argument highlights a gap in currently accepted models of phonologization such as Ohala (1989, 1992, 1993), on which Hale & Reiss base their criticism of OT. In these models, phonologization is driven by misparsing, which is in turn caused by the inherent ambiguity of local phonetic data.[34] From this viewpoint, affrication and spirantization occur because hearers can misinterpret an excessively noisy stop release as the realization of a fricative target or because articulatory undershoot on the part of the speaker can compromise the perceptibility of the oral closure. Notice, however, that both release noise and closure duration are purely local phonetic properties of plosives; they are independent of the segmental inventory of the language, except insofar as the need to preserve phonemic contrasts limits the average range of phonetic realizations for particular segments. In this sense, purely bottom-up approaches to phonologization fail to provide a mechanism for adjusting the relative probability of phonologization events in the light of their global systemic effects (see Bermúdez-Otero & Hogg, 2003: §1.1).

Optimality-theoretic constraints afford precisely such a mechanism, as they capture not only the local but also the global implications of markedness relationships. Assume, for example, that infants have access to a constraint set that includes VOP, *[-son, +cont], and IDENT-voice,[35] but – crucially – does not include a constraint against voiceless obstruents. In such a situation, the infant cannot construct a grammar that generates fricatives or voiced plosives while simultaneously forbidding voiceless stops; this avenue is closed even if, by virtue of their local phonetic properties, the voiceless plosives that occur in the ambient language invite alternative phonological categorizations. Note, however, that patterns established by mere induction over input data would not be subject to such restrictions. Thus, the constraint set available to the infant embodies the knowledge that voiceless stops must be grammatical; such knowledge transcends mere inductive generalization.

6.5. Markedness constrains grammatical restructuring

The previous section showed that linguistic typology emerges from the interaction between performance phenomena and markedness constraints: whilst phonologization and grammaticalization routinely filter marked structures out of language, cognitive representations of markedness (in the form of optimality-theoretic constraints) block performance-driven changes with impermissible global effects. In the second case, constraints are causally efficient and cannot therefore be dismissed as epiphenomenal. In this section we will show that markedness constraints also set limits to diachronic processes of grammatical restructuring (cf. §5.3). Insofar as these processes are driven by internal factors,

the involvement of markedness constraints cannot be a mere epiphenomenon of mechanisms external to the grammar.

To prove this point, we will use a test-case favoured by the critics of OT: rule inversion (§5.3).[36] It will be recalled that, according to the critics of OT, hiatus-filling processes arising from the inversion of coda deletion rules are arbitrary: the segment inserted in the hiatus site will be identical with whatever consonant happened to be targeted by the antecedent rule of coda deletion; the hiatus-filler may therefore be a highly marked segment, in violation of the emergence of the unmarked (§4.3). Against such claims, we shall demonstrate that hiatus rules arising by inversion are subject to ordinary markedness restrictions.

Consider the phenomenon of *l*-intrusion found in certain dialects of English spoken in the Northeast of the United States (Gick, 1999, personal communication). In a former period of the history of these dialects – call it 'Stage I' – underlying /l/ surfaced in the onset but was subject to vocalization in the coda. This gave rise to a pattern of alternations in stem- and word-final position known as 'linking *l*':

(26)   **Linking *l***

	a.		b.	
	draw	*drawing*	*drawl*	*drawling*
UR	/dɹɔː/	/dɹɔː-ɪŋ/	/dɹɔːl/	/dɹɔːl-ɪŋ/
SR	[dɹɔː]	[dɹɔː.ɪŋ]	[dɹɔː]	[dɹɔː.lɪŋ]

Subsequently, alternations such as (26b) were reanalysed: the underlying representation of the stem was restructured in the image of the preconsonantal/prepausal alternants,[37] and the [l] found in prevocalic environments was reinterpreted as an epenthetic hiatus-filler. Concomitantly, lexical distinctions such as /dɹɔː/ vs /dɹɔːl/ collapsed, giving rise to a new pattern of alternations known as 'intrusive *l*':

(27)   **Intrusive *l***

	a.		b.	
	draw	*drawing*	*drawl*	*drawling*
UR	/dɹɔː/	/dɹɔː-ɪŋ/	/dɹɔː/	/dɹɔː-ɪŋ/
SR	[dɹɔː]	[dɹɔː.lɪŋ]	[dɹɔː]	[dɹɔː.lɪŋ]

Example (27a) shows that the alternation between [l] and ∅ underwent reanalysis after /ɔː/. There is, however, a problem: at Stage I there were also [l]~∅ alternations after other vowels, including /ɑː/ and /ə/.

(28)   **Linking *l* after /ə/**

	cruel	*cruel act*
UR	/kɹuːəl/	/kɹuːəl ækt/
SR	[kɹuː.wə]	[kɹuː.wə.lækt]

This is important because, like /ɔː/, the vowels /ɑː/ and /ə/ do not have high offglides and are allowed in word-final position; they are therefore capable of occurring in hiatus environments (McCarthy, 1993: 176; Gick, 1999: 37). By the same token, the alternation between [l] and ∅ after /ɑː/ and /ə/ could be reanalysed in the same way as after /ɔː/, with [l] being interpreted as an epenthetic hiatus-filler. This would give rise to *l*-intrusion after /ɑː/ and /ə/. Yet, intriguingly, this reanalysis scarcely happened: in the vast majority of dialects there is *l*-intrusion after /ɔː/ but not after /ɑː/ and /ə/.

(29)   **The failure of *l*-intrusion after /ɑː/ and /ə/**

   law[l]-*abiding*   but   *the bra*[∅] *is*

The facts demand an explanation, but it is clear that an appeal to mere inductive generalization over the primary data available to infants at Stage I must necessarily fail: the alternations in (26b) and (28) are analogous in every respect, and so both should be equally liable to reanalysis.

    The key to the problem lies in the fact that, in the relevant dialects, /l/ and /ɔː/ have identical feature specifications under the V-Place node. This is confirmed by phonetic evidence showing that both segments are articulated with identical gestures of tongue dorsum backing, creating similar vocal tract configurations in the pharyngeal and uvular regions (Gick, Kang & Whalen, 2002). In other words, by subtracting the central coronal closure from /l/ one obtains /ɔː/. In contrast, /ɑː/ and /ə/ have different V-Place features from /l/.

    In this light, it becomes clear that, at the end of Stage I, infants reinterpreted linking [l] as the least marked sonorant that was *both* capable of filling an empty onset *and* homorganic with a preceding [ɔː]. Linking [l] after [ɑː] and [ə] was not so reanalysed because these vowels possess different V-Place specifications. Tableau (30) illustrates the workings of the grammar that resulted from this reanalysis. The following constraints are involved: $*G_{[-hi]}$ prohibits glides (i.e. nonnuclear vocoids) with nonhigh V-Place specifications (Uffmann, 2002); DEP-VPlace is violated by the insertion of V-Place features; DEP-[−son] blocks obstruent insertion; ONSET requires that syllables should have an onset; DEP-CPlace militates against the insertion of consonantal place features; and, finally, *[lateral] asserts that lateral consonants are marked.

[INSERT (30) HERE]

    The significance of the partial reanalysis of linking *l* should now be clear. Infants did not expect [l] to behave as a hiatus-filler independently of the melodic context; rather, they accepted intrusive [l] only when homorganic with the preceding vowel, where it could be construed as an alternative to a nonhigh glide – nonhigh glides being highly marked and absent from the segment inventory of English. Crucially, infants rejected *l*-intrusion after /ɑː/ and /ə/ because, in that environment, [l]-epenthesis would be no more faithful than [j]- or [w]-insertion[38] but would in addition incur an extra violation of *[lateral]. Thus, learners tackled the analysis of alternations such as (26b) and (28) armed with three pieces of knowledge:

(i)   knowledge that epenthetic segments should be unmarked (in compliance with the emergence of the unmarked; see §4.3);

(ii)   knowledge that, *ceteris paribus*, laterals are marked;

(iii)   knowledge that nonhigh glides are marked.

Of these three items of knowledge, only (iii) could be arrived at inductively, since nonhigh glides were absent from the primary linguistic data. In contrast, (i) and (ii) clearly lay beyond the reach of inductive generalization.

Thus, the evidence of *l*-intrusion does not contradict – but actually supports – OT's claim that the possible outcomes of grammatical restructuring are constrained by markedness. Note, incidentally, that our analysis in no way predicts that *l*-intrusion can arise historically by simple phonologization, instead of rule inversion (cf. Blevins, to appear: note 13). Given a statement about the universal constraint set, one can use factorial typology to deduce purely synchronic predictions about possible and impossible grammars; these synchronic predictions have no diachronic implications beyond the simple – though crucial – fact that historical change cannot create an impossible language. In this sense, the causes of diachronic constraint reranking fall beyond the purview of the theory of CON. Accordingly, the constraint set postulated in tableau (30) makes predictions about what processes of [l]-epenthesis can or cannot occur in natural language (see e.g. note 38); it further entails that no historical change resulting in an impermissible [l]-epenthesis process may occur; but it tells us nothing about how permissible [l]-epenthesis processes arise diachronically.

To conclude, in §6.4 and §6.5 we have demonstrated that knowledge of markedness plays an essential rôle in constraining both performance-driven change and grammatical restructuring. It follows that children do rely on cognitive representations of markedness (in the form of optimality-theoretic constraints) to transcend the limits of induction over primary linguistic data. Obviously, we do not wish to deny that inductive generalization plays a significant rôle in language acquisition; we suspect, in fact, that it may be central to the acquisition of much morphology. Induction cannot however bear the heavy burden that Hale & Reiss choose to impose on it (§6.1). Notably, there exist markedness-governed linguistic patterns that cannot be acquired by mere inductive generalization over a corpus of utterances, even if UG supplies a representational format and a simplicity measure for linguistic rules (§6.4).

This concludes our demonstration that, despite their nonarbitrary character, optimality-theoretic constraints do not violate Ockham's Razor. On these grounds we reject the charge of 'substance abuse'.

7. The origins of grounded constraints

Previous sections have shown that there are markedness-governed linguistic patterns that cannot be acquired by mere inductive generalization over a corpus of utterances. During language acquisition, therefore, children must have some alternative means of access to knowledge of markedness. Such knowledge, we assume, takes the form of optimality-theoretic constraints. The question thus remains: if markedness constraints do not arise by induction over primary linguistic data, what is their origin and how do they develop their nonarbitrary relationship with extragrammatical substance?

There are, in principle, two possible accounts of the origin of markedness constraints. In a phylogenetic account, constraints are supplied innately by UG. Subject to nonpathological maturation, constraints become available to the infant in the absence of all experience. In this context, explaining the nonarbitrary character of constraints is a task for evolutionary biology. In an ontogenetic account, in contrast, constraints arise developmentally through the dynamic interaction between the grammar and related performance systems. In this account, the emergence of constraints requires experience; the

relevant input, however, does not consist of a corpus of utterances, but rather of the child's active manipulation of the cognitive and physiological systems with which the grammar interacts: e.g. the conceptual-intentional system, the parser, systems of motor control, etc. From this viewpoint, accounts of grounding are to be sought in developmental cognitive psychology.

As we show in §7.1, the hypothesis that CON is innate faces very severe challenges. In this light, we will argue that linguists should be prepared to countenance ontogenetic approaches to the origin of markedness constraints; in §7.2 we survey some possible scenarios.

7.1. Are constraints innate?

In the optimality-theoretic literature it has often been assumed that the composition of the universal constraint set is determined by UG alone: all constraints are innate and available to the infant in the absence of all experience. However, the innateness hypothesis faces two obstacles: it fails to provide credible accounts of either the epigenesis or the phylogenesis of CON.

The epigenesis of CON is the developmental process whereby neural representations of constraints are built in individual brains under the direction of the genome. We assume that the representation of constraints in the brain involves fine-grained patterns of cortical connectivity (Elman *et al.*, 1996). Under the innateness hypothesis, the genes fully control the development of such patterns. This claim is problematic for two reasons. First, a full specification of all relevant neuronal structures would require vast amounts of genetic information (see Marcus, 2001: 155-6, for germane comments). The amount of genetic information required might conceivably be minimized if the epigenesis of constraints were largely guided by purely physical 'laws of form' (Thompson, 1942), of the sort that Minimalist syntacticians portentously – though rather obscurely – invoke (Uriagereka, 1998; Chomsky, 2001; Freidin & Vergnaud, 2001); but, from a strictly formal viewpoint, constraints are far too complex, irregular, and unpredictable for that. Secondly, the hypothesis of full genetic control clashes with empirical evidence that cortical development is highly flexible and input-dependent (see Elman *et al.*, 1996: ch. 5). Marcus (2001: §6.3) qualifies these observations with arguments that show that the general neural machinery needed for symbol-manipulation can in fact be innate; the hypothesis that all members of CON are genetically specified, however, far exceeds Marcus's claims, as it concerns a large set of complex domain-specific symbolic representations.

The phylogenesis of CON is the evolutionary process whereby the genes responsible for directing the epigenesis of CON became a component of the *homo sapiens* genotype. Under the innateness hypothesis, evolutionary mechanisms assembled a complex of genes capable of installing representations of every markedness constraint in each phenotype. This claim is again implausible for two reasons:

First, one could argue that the dominant evolutionary mechanism in the phylogenesis of CON is exaptation (Gould & Lewontin, 1979; Gould & Vrba, 1982; Gould, 1991). In this view, CON is a 'spandrel': it originally evolved either as an adaptation for a purpose other than linguistic computation or as a mere nonfunctional by-product of something else, and it was subsequently coopted for linguistic purposes. However, CON lacks the characteristic properties of spandrels. As we noted above with regard to the 'laws of form', CON is far too complex and rich to arise as a mere by-product of some other adaptive feature under general design constraints. Similarly, if it constitutes the redeployment of an initially adaptive character for a new purpose, the original function of this character remains to be determined. In addition, spandrels typically possess properties that seem arbitrary from the viewpoint of optimal design; but constraints hug the nonlinguistic ground rather closely (*modulo*

simplicity, self-containment, and language-specific relationships of strict dominance; see §6.2 and §6.3).

Secondly, one might suggest that it was adaptation, rather than exaptation, that played the central rôle in the phylogenesis of CON; but this scenario is equally implausible. In essence, for each markedness constraint M it would have to be shown that an individual possessing genes that resulted in the hard-wiring of M enjoyed some degree of reproductive advantage over another individual lacking those genes. It is clear that, in many circumstances, such an advantage would simply not exist: for speakers of a CVC language, for example, being innately endowed with constraints on the composition of complex onsets would serve no practical purpose (Hayes, 1999a: §13). Even where a reproductive advantage existed, it would still take enormously long time-spans for natural selection to produce its effects. See moreover McMahon (2000: ch. 5) for much relevant discussion.

We therefore conclude that UG is unlikely to supply all – or even most – markedness constraints. Admittedly, the arguments that we have deployed in this section are of a rather general and imprecise nature. This is to some extent unavoidable in the current state of knowledge. In our view, however, our discussion suffices to shift the burden of proof against the innateness hypothesis if plausible alternative accounts of the origin of constraints become available (see §7.2). Nonetheless, there is a danger that, owing to their generality and imprecision, arguments of this kind may be used indiscriminately against any form of linguistic innateness. In this connection, we do not wish to deny that important components of linguistic knowledge may be innate; in fact, a vast amount of grammatical structure can be innate even if (most) constraints are not. Notably, we have repeatedly stressed that constraints are stated in terms of specifically linguistic categories; we will not here venture to suggest which of these categories may be innate and which may emerge from developmental processes such as we postulate in §7.2.[39] Similarly, we are willing to countenance the possibility that (aspects of) the representational format in which linguistic expressions are made available to central cognitive systems may be innate. Moreover, it is likely that the general architecture of grammar is also innate or develops fairly directly from innate neural biases. In this context, therefore, denying that all markedness constraints are innate is by no means equivalent to postulating a linguistic *tabula rasa*.

7.2. Grounded constraints as ontogenetically emergent

The arguments outlined in §7.1 invite us to consider the possibility that CON may to a large extent not be innate. We know, however, that CON cannot be discovered by inductive generalization over input data, for infants depend upon it to overcome the limitations of induction. We are therefore forced to pursue a third hypothesis: language learners acquire constraints on the basis of experience, but the relevant experience does not consist of exposure to a series of utterances in the ambient language; it involves, rather, the child's monitoring of her own linguistic performance. Such an account depicts constraints as *ontogenetically emergent* and emphasizes the dynamic interaction between grammar and performance systems in the process of linguistic development.[40] In this section we review some of the research in which this avenue is beginning to be explored.

Hayes (1999a) proposes an ontogenetic approach to the origin of phonological markedness constraints. Setting aside for the moment the task of assigning underlying phonological representations to lexical items, Hayes regards the construction of a phonological grammar as involving two (perhaps overlapping) stages: constraint discovery and constraint ranking. He assumes that constraints are ranked by means of an error-driven demotion algorithm (see e.g. Tesar & Smolensky, 2000: ch. 3). For the logically prior

operation of constraint discovery, Hayes proposes that children rely on three innately specified cognitive devices:

(31)   **Devices for phonological constraint discovery in Hayes (1999a)**
   a.   generator of phonetic difficulty landscapes
   b.   constraint generator
   c.   constraint evaluator

First, the child uses her own experience of articulation and perception to assess the relative performance difficulty of different phonological structures; on this basis, she constructs phonetic difficulty landscapes analogous to that in (22) above. Since, in nonpathological cases, the physical and physiological factors that determine performance difficulty are identical from child to child, the difficulty landscapes created by different children will be nondistinct. In addition, Hayes postulates a computational system which, given the primitive elements of phonological representation, generates all conceivable constraint statements. The candidates emitted by this constraint generator are then evaluated according to their degree of grounding: the phonetic effectiveness of each candidate is measured by checking its predictions of relative difficulty against the phonetic landscape; a candidate is selected if it has greater phonetic effectiveness than all minimally different constraints of equal or lesser specificity (for details, see Hayes, 1999a: §8, §9). Thus, successful constraints are 'local maxima of good phonetic fit'.

   Hayes is primarily concerned with striking the right balance between phonetic effectiveness and formal simplicity (see §6.2): by stipulating that each possible constraint competes only with candidates that are equally or more general, Hayes ensures that a large number of relatively simple constraints will be selected even if their phonetic accuracy is not perfect. The theory that Hayes devises in pursuance of this goal is admirable for its explicitness and precision and, consequently, for its empirical falsifiability. From an ontological viewpoint, however, we doubt whether Hayes's proposals can be accorded a realist interpretation:

   First, the innate devices listed in (31) are considerably complex; accounting for their epigenesis and phylogenesis (see §7.1) therefore poses a serious challenge. The problem is particularly acute, we feel, in the case of the proposed constraint generator and constraint evaluator. For Hayes's generator of phonetic difficulty landscapes, in contrast, there is some independent empirical support: e.g. Steriade (2001: 235-9) adduces evidence from poetic rhyming and loan adaptation to show that speakers possess 'a model of the generic listener's perceptual abilities and biases'; more specifically, this model or 'P-map' encodes knowledge about the relative discriminability of sound contrasts.

   Secondly, Hayes's strategy of profuse constraint generation followed by selection adds to the problem of efficient computation that already affects standard OT. The definition of the constraint discovery system *qua* function is not compromised by Hayes's assumption that infinite constraint candidates are first generated and then assessed. Nonetheless, designing an efficient computational algorithm for such a function raises nontrivial problems. (For similar observations in respect of the grammar itself, see McCarthy, 2002: §1.1.3, §4.3; Tesar & Smolensky, 2000: ch. 8.; Kuhn, 2003.)

   Despite Steriade's support for the 'P-map', some of the problems of Hayes's theory could be avoided by allowing constraints to emerge in direct response to performance self-monitoring, without mediation from difficulty landscapes. Thus, following Boersma (1998b),[41] we may assume that the learner posits a high-ranking markedness constraint M against grammatical output representations containing a structure *s* whenever *s* fails to be

successfully interpreted by an appropriate performance system. Consider, for example, a child acquiring a language such as Russian, which tolerates heterorganic N.C clusters: e.g. [anglijə] 'England'. Suppose that, upon first exposure to words containing such clusters, the child stores them in her protolexicon in something approaching the adult phonological output representation: i.e. /anglijə/ (see Tesar & Smolensky, 2000: 76, and references therein). When these representations are submitted for interpretation by the child's articulatory system, however, problems are likely to occur: e.g. because the child lacks the necessary motor control skills, the velar gesture associated with /g/ is likely to overlap with the articulation of the preceding nasal, yielding an utterance perceived as *[aŋglijə]. Such instances of unsuccessful performance alert the child to the inherent difficulty of realizing coda-onset clusters in which the coda licenses its own place features. In response, the child inserts a constraint such as CODACOND-Place (see (1) above) at the top of her emergent constraint hierarchy. Repeated exposure to items such as [anglijə], however, will prompt the child to continue to explore her articulatory capabilities, until she eventually arrives at a correctly timed gestural score for heterorganic N.C clusters. At this point, phonological output representations containing structures such as [-n.g-] become interpretable; concomitantly, CODACOND-Place is demoted below IDENT-Place.

Boersma's proposals have a number of interesting implications. In Hayes's theory, for example, the process of constraint discovery yields an unordered constraint set, which is then subject to ranking. To avoid the Subset Problem, however, the constraint set has to be initially divided into two strata – with all markedness constraints dominating all faithfulness constraints – before being submitted to the demotion algorithm; see footnote 28 above and Tesar & Smolensky (2000: 76). In contrast, Boersma's theory does not separate the process of constraint discovery from that of ranking; rather, markedness constraints enter the hierarchy directly from above and are progressively demoted. Accordingly, the initial state of the constraint hierarchy need not be stipulated.

Boersma's key insights can also be articulated in ways that address Hayes's (1999a) concern with the relative simplicity of constraint statements (see above and §6.2). One possibility is that, at each point in linguistic development, the child formulates constraints against currently uninterpretable output configurations in the most general terms possible. Consider, for example, the developmental path of an infant acquiring a language with a phoneme inventory that includes both short and long voiced plosives: e.g. /b, d, g, bː, dː, gː/ (see §6.1, §6.2). At an early stage, any of these segments is likely to raise phonetic difficulties for the child. If the child responds to such difficulties by formulating the most inclusive constraint possible, then VOP – which prohibits *all* voiced obstruents – will be inserted at the top of the hierarchy. Subsequently, however, the child will master the production of some of the short voiced plosives; at this point, VOP will be demoted, but lingering difficulties with /bː, dː, gː/ will cause a more specific constraint VOP-long to enter the grammar from above. Interestingly, it follows that, in such cases, the more specific constraint (e.g. VOP-long) initially dominates the more general constraint (e.g. VOP); this is a welcome prediction because the more specific constraint is totally redundant under the opposite ranking (see McCarthy, 2002: 44, for references to the literature on stringency relationships among constraints).

More generally, Boersma's approach implies that the knowledge of markedness eventually attained by each child is bound by her experience of linguistic performance. Each infant may thus be expected to explore only a subregion of the space of possible output candidates for each input (crucially including the grammatical output), and to posit only that subset of markedness constraints that is relevant to the structures of her language. Under this

hypothesis, a great deal of the machinery associated with GEN and CON in standard OT may exist only *virtually* in individual grammars (Boersma, 1998b: §5).[42] We may therefore hope that a theory of linguistic ontogenesis developed along these lines will eventually lead to a much leaner and more streamlined characterization of the adult speaker's mature competence.

At present, this approach to the origin of markedness constraints remains at the programmatic stage; time will tell how far it can develop. We nonetheless feel that it deserves serious attention from phonologists and syntacticians. As we saw in §6 (specially §6.1), the logical problem of language acquisition is all too often framed as a simple choice between innateness and induction over input data. By neglecting performance self-monitoring as a potential source of linguistic knowledge, however, this style of thought clouds the issues. Notably, it makes it almost unintelligible that linguistic knowledge should at once transcend the limits of inductive generalization and display close functional adaptation (see §7.1 for the problems of evolutionary approaches to this fact, though cf. Pinker & Bloom, 1990; Newmeyer, 1991). Yet, as our discussion has shown, markedness constraints characteristically exhibit both properties. Failing to understand the nature of markedness effects may however have disastrous consequences: if our arguments in §3, §4, and §5 are correct, markedness plays a pervasive rôle in both phonology and syntax.

Admittedly, the idea that markedness constraints are ontogenetically emergent arose first in phonology. We would not however wish to attach any particular significance to this fact. In this connection, our discussion of Boersma's key insights is deliberately framed in terms that apply both to phonology and to syntax. Again, our comparison of the two in sections §3 to §5 does not give grounds to expect that markedness effects in syntax should be essentially different from those in phonology.

8. Conclusion

Plato's Problem (Chomsky, 1986) poses the formidable task of discovering the origins of linguistic knowledge. Faced with this challenge, orthodox generative theory, as articulated in Hale & Reiss's and Newmeyer's critiques of OT, contemplates a stark choice. Linguistic knowledge – it suggests – has only two sources. There is, on the one hand, the innately specified knowledge supplied by UG. This is either decidedly 'substance-free' (Hale & Reiss) or, at best, arbitrary but for the rather limited rôle that natural selection may have played in its hard-wiring (§7.1; though cf. Pinker & Bloom, 1990; Newmeyer, 1991). On the other hand, there is inductive generalization over linguistic expressions, operating through the hallowed empiricist mechanisms of analogy and association. In the main, linguistic patterns that reveal close functional adaptation arise through induction over linguistic data that have been sifted through the Arena of Use – or so the argument goes.

This, we have argued, is a false dichotomy. As has been emphasized in the neuroconstructivist tradition, knowledge of language can also emerge developmentally from the learner's dynamic interaction with her environment, with innate specifications[43] regulating – rather than determining – the ontogenetic pathway (Karmiloff-Smith, 1998). Where all the relevant factors affecting development are universal, the emerging knowledge will be universal too. If the proposals surveyed in §7.2 are on the right track, this is precisely the source of the knowledge of markedness embodied in many optimality-theoretic constraints. Crucially, as was shown in §6, such knowledge transcends the limitations of inductive generalization; despite Hale & Reiss's strictures, therefore, there is no violation of Ockham's Razor involved in according cognitive reality to markedness patterns grounded in nonlinguistic substance. *Pace* the Representational Hypothesis, moreover, we have argued that markedness constraints are likely to play as essential a rôle in syntax as in phonology (§4, §5).

Finally, it has been one of our main goals in this article to elucidate the ontology of optimality-theoretic constraints. Our discussion has identified some of their key attributes and shown how these properties can coexist. Markedness constraints, we suggest, are self-contained but partially nonarbitrary and, whilst universal, are likely to be developmentally emergent.

REFERENCES

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

Ackema, P., Neeleman, A., 2000. Absolute ungrammaticality. In: Dekkers, van der Leeuw, van de Weijer (Eds.), pp. 279-301.

Aissen, J., 1999. Markedness and subject choice in Optimality Theory. Natural Language and Linguistic Theory 17, 673-711.

Aissen, J., 2001. Markedness and subject choice in Optimality Theory. In: Legendre, Grismshaw, Vikner (Eds.), pp. 61–96.

Anderson, J.M., 1986. Structural analogy and case grammar. Lingua 70, 79-129.

Anderson, J.M., this volume. Structural analogy and universal grammar.

Anderson, J.M., Ewen, C., 1987. Principles of Dependency Phonology. Cambridge University Press, Cambridge.

Anderson, S.R., 1977. On mechanisms by which languages become ergative. In: Li, C.N. (Ed.), Mechanisms of syntactic change. University of Texas Press, Austin, pp. 317-63.

Anderson, S.R., 1981. Why phonology isn't "natural". Linguistic Inquiry 12, 493-539.

Anderson, S.R., 1992. A-morphous morphology. Cambridge University Press, Cambridge.

Anderson, S.R., 1996. How to put your clitics in their place; or, why the best account of second-position phenomena may be something like the optimal one. The Linguistic Review 13, 165-191.

Anderson, S.R., 2000. Towards an optimal account of second position phenomena. In: Dekkers, van der Leeuw, van de Weijer (Eds.), pp. 302-33.

Archangeli, D., Langendoen, T. (Eds.), 1997. Optimality Theory: an overview. Blackwell, Oxford.

Archangeli, D., Pulleyblank, D., 1994. Grounded phonology. The MIT Press, Cambridge, MA.

Bach, E., Harms, R.T., 1972. How do languages get crazy rules? In: Stockwell, R.P., Macaulay, R.K.S. (Eds.), Linguistic change and generative theory: Essays from the UCLA Conference on Historical Linguistics in the Perspective of Transformational Theory, February 1969. Indiana University Press, Bloomington, IA, pp. 1-21.

Baković, E., Keer, E., 2001. Optionality and ineffability. In: Legendre, Grimshaw, Vikner (Eds.), pp. 97-112.

Barbosa, P., Fox, D., Hagstrom, P., McGinnis, M., Pesetsky, D. (Eds.), 1998. Is the best good enough? Optimality and competition in syntax. MIT Press, Cambridge, MA.

Bat-El, O., 1996. Selecting the best of the worst: the grammar of Hebrew blends. Phonology 13, 283-328.

Beckman, J., Dickey, L.W., Urbanczyk, S. (Eds.), 1995. University of Massachusetts Occasional Papers in Linguistics, 18: Papers in Optimality Theory. Graduate Linguistic Student Association, Amherst, MA.

Beddor, P.S., Krakow, R.A., Lindemann, S., 2001. Patterns of perceptual compensation and their phonological consequences. In: Hume, Johnson (Eds.), pp. 55-78.

Bermúdez-Otero, R., 1998. Prosodic optimization: the Middle English length adjustment. English Language and Linguistics 2, 169-197.

Bermúdez-Otero, R., 1999. Constraint interaction in language change [Opacity and globality in phonological change]. PhD dissertation, University of Manchester & Universidad de Santiago de Compostela.

Bermúdez-Otero, R., 2002. Review of Burton-Roberts, Carr, Docherty (Eds.), 2000. Journal of Linguistics 38, 403-410.

Bermúdez-Otero, R., Hogg, R.M., 2003. The actuation problem in Optimality Theory: phonologization, rule inversion, and rule loss. In: Holt, D. E. (Ed.), Optimality Theory and language change. Kluwer, Dordrecht, pp. 91-119.

Blevins, J., 1997. Rules in Optimality Theory: two case studies. In: Roca (Ed.), pp. 227-260.

Blevins, J., to appear. Consonant epenthesis: natural and unnatural histories. In: Good, J. (Ed.), Proceedings of the Workshop on Explaining Linguistic Universals. Oxford University Press, Oxford.

Boersma, P., 1997. How we learn variation, optionality, and probability. Proceedings of the Institute of Phonetic Sciences of the University of Amsterdam 21, 43-58.

Boersma, P., 1998a. Functional phonology: formalizing the interaction between articulatory and perceptual drives. Holland Academic Graphics, The Hague.

Boersma, P., 1998b. Typology and acquisition in functional and arbitrary phonology. Paper presented at the Utrecht Phonology Workshop, 22 June 1998. Available at http://www.fon.hum.uva.nl/paul/papers/typ_acq.pdf.

Boersma, P., Dekkers, J., Weijer, J. van de, 2000. Introduction – Optimality Theory: phonology, syntax, and acquisition. In: Dekkers, van der Leeuw, van de Weijer (Eds.), pp. 1-44.

Boersma, P., Hayes, B., 2001. Empirical tests of the gradual learning algorithm. Linguistic Inquiry 32, 45-86.

Börjars, K., Donohue., M., 2000. Much ado about nothing: features and zeroes in Germanic noun phrases. Studia Linguistica 54, 309-53.

Börjars, K., Vincent, N., 2000. Multiple case and the 'wimpiness' of morphology. In: Butt, M., King, T. H. (Eds.), Argument realization. CSLI Publications, Stanford, CA, pp. 15-40.

Bresnan, J., 2000a. Optimal syntax. In: Dekkers, van der Leeuw, van de Weijer (Eds.), pp. 334-385.

Bresnan, J., 2000b. Pidgin genesis and Optimality Theory. In: Siegel, J. (Ed.), Processes of language contact: Case studies from Australia and the Pacific. Les Editions Fides, Montreal, pp. 145-73.

Bresnan, J., 2001a. Lexical-functional syntax. Blackwell, Oxford.

Bresnan, J., 2001b. The emergence of the unmarked pronoun: Chicheŵa pronominals in Optimality Theory. In: Legendre, Grimshaw, Vikner (Eds.), pp. 113-142.

Bresnan, J., Aissen, J., 2002. Optimality and functionality: objections and refutations. Natural Language and Linguistic Theory 20, 81-95.

Bresnan, J., Dingare, S., Manning, C., 2001. Soft constraints mirror hard constraints: voice and person in English and Lummi. In: Butt, M., King, T.H. (Eds.), Proceedings of the LFG 01 Conference. University of Hong Kong, on-line proceedings, CSLI Publications, Stanford: http://csli-publications.stanford.edu/.

Brody, M., 1995. Lexico-logical form: a radically minimalist theory. MIT Press, Cambridge, MA.

Burton-Roberts, N., 2000. Where and what is phonology? A representational perspective. In: Burton-Roberts, Carr, Docherty (Eds.), pp. 39-66.

Burton-Roberts, N., Carr, P., Docherty, G. (Eds.), 2000. Phonological knowledge: conceptual and empirical issues. Oxford University Press: Oxford.

Burton-Roberts, N., Poole, G., this volume. Syntax vs phonology: a representational approach to stylistic fronting and verb-second in Icelandic.

Burzio, L., 1998. Anaphora and soft constraints. In: Barbosa, Fox, Hagstrom, McGinnis, Pesetsky (Eds.), pp. 93-113.

Carr, P., 2000. Scientific realism, sociophonetic variation, and innate endowments in phonology. In: Burton-Roberts, Carr, Docherty (Eds.), pp. 67-104.

Carr, P., this volume. Universal Grammar and syntax/phonology parallelisms.

Carstairs-McCarthy, A., 1999. The origins of complex language: An inquiry into the evolutionary beginnings of sentences, syllables and truth. Oxford University Press, Oxford.

Chomsky, N., 1986. Knowledge of language: its nature, origin, and use. Praeger, Westport, Conn.

Chomsky, N., 1995. The Minimalist Program. MIT Press, Cambridge, MA.

Chomsky, N., 2001. Beyond explanatory adequacy. MIT Occasional Papers in Linguistics 20.

Chomsky, N., with Belletti, A., Rizzi, L., 2002. An interview on minimalism. In: Chomsky, N. (ed. by Belletti, A., Rizzi, L.), On nature and language. Cambridge University Press, Cambridge, pp. 92-161.

Chomsky, N., Halle, M., 1968. The sound pattern of English. Harper and Row, New York.

Chung, S., 1978. Case marking and grammatical relations in Polynesian. University of Texas Press, Austin.

Clements, G.N., 1990. The role of the sonority cycle in core syllabification. In: Kingston, J., Beckman, M. (Eds.), Papers in laboratory phonology, 1: Between the grammar and physics of speech. Cambridge University Press, Cambridge, pp. 283-333.

Clements, G.N., 1992. The sonority cycle and syllable organization. In: Dressler, W.U., Luschützky, H.C., Pfeiffer, O.E., Rennison, J.R. (Eds.), Phonologica 1988: Proceedings of the Sixth International Phonology Meeting. Cambridge University Press, Cambridge, pp. 63-76.

Cole, J., Iskarous, K., 2001. Effects of vowel context on consonant place identification: implications for a theory of phonologization. In: Hume, Johnson (Eds.), pp. 103-22.

Cole, M., 2000. The syntax, morphology and semantics of null subjects. Unpublished PhD thesis, University of Manchester, Manchester.

Costa, J., 2001. The emergence of the unmarked word order. In: Legendre, Grismshaw, Vikner (Eds.), pp. 171-203.

Croft, W., 1995. Autonomy and functionalist linguistics. Language 71, 490-532.

Croft, W., 2001. Radical construction grammar: syntactic theory in typological perspective. Oxford University Press, Oxford.

Davis, S., 2000. Some analytical issues in Optimality Theory. In: Ritter (Ed.), pp. 117-133.

DeLancey, S., 1981. An interpretation of split-ergativity and related patterns. Language 57, 626-57.

Dekkers, J., Leeuw, F. van der, Weijer, J. van de (Eds.), 2000. Optimality Theory: phonology, syntax, and acquisition. Oxford University Press, Oxford.

Du Bois, J. W., 1987. Discourse basis of ergativity. Language 63, 805-55.

Elman, J.L., Bates, E.A., Johnson, M.H., Karmiloff-Smith, A., Parisi, D., Plunkett, K., 1996. Rethinking innateness: a connectionist perspective on development. The MIT Press, Cambridge, MA.

Evans, N., 1995. Current issues in the phonology of Australian languages. In: Goldsmith, J. A. (Ed.), pp. 723-61.

Freidin, R., Vergnaud, J.-R., 2001. Exquisite connections: some remarks on the evolution of linguistic theory. Lingua 111, 639-666.

García, E., 1979. Discourse without syntax. In: Givón, T. (Ed.), Discourse and syntax. Academic Press, New York, pp. 23-49.

Gerfen, C., 2001. A critical view of licensing by cue: codas and obstruents in Eastern Andalusian Spanish. In: Lombardi, L. (Ed.), Segmental phonology in Optimality Theory: constraints and representations. Cambridge University Press, Cambridge, pp. 183-205.

Gick, B., 1999. A gesture-based account of intrusive consonants in English. Phonology 16, 29-54.

Gick, B., Kang, A M., Whalen, D H., 2002. MRI evidence for commonality in the post-oral articulations of English vowels and liquids. Journal of Phonetics 30, 357-372.

Gilligan, G M., 1987. A cross-linguistic approach to the pro-drop parameter. Unpublished PhD dissertation, University of Southern California.

Givón, T., 1995. Functionalism and grammar. Benjamins, Amsterdam.

Goldsmith, J.A. (Ed.), 1995. The handbook of phonological theory. Blackwell Publishers, Cambridge, MA.

Gould, S.J., 1991. Exaptation: a crucial tool for an evolutionary psychology. Journal of Social Issues 47, 43-65.

Gould, S.J., Lewontin, R.C., 1979. The spandrels of San Marco and the Panglossian paradigm: a critique of the adaptationist programme. Proceedings of the Royal Society of London B 205, 581-598.

Gould, S.J., Vrba, E.S., 1982. Exaptation – a missing term in the science of form. Paleobiology 8, 4-15.

Grimshaw, J., 1997. Projection, heads, and optimality. Linguistic Inquiry 28, 373-422.

Groat, E., O'Neil, J., 1996. Spell-Out at the LF interface. In: Abraham, W., Epstein, S. D., Thráinsson, H., Zwart, C.J.-W. (Eds.), Minimal ideas: syntactic studies in the minimalist framework. John Benjamins, Amsterdam, pp. 113-139.

Haiman, J., 1985. Natural syntax: iconicity and erosion. Cambridge University Press, Cambridge.

Hale, M., 2000. Marshallese phonology, the phonetics-phonology interface and historical linguistics. In: Ritter (Ed.), pp. 241-257.

Hale, M., Reiss, C., 2000. Phonology as cognition. In: Burton-Roberts, Carr, Docherty (Eds.), pp. 161-184.

Haspelmath, M., 1999. Optimality and diachronic adaptation. Zeitschrift für Sprachwissenschaft 18, 180-205.

Hawkins, J.A., 1989. Competence and performance in the explanation of language universals. In: Arnold, D., Atkinson, M., Durand, J., Grover, C., Sadler, L. (Eds.), Essays on grammatical theory and Universal Grammar. Clarendon Press, Oxford, pp. 119-152.

Hawkins, J.A., 2001. Why are categories adjacent. Journal of Linguistics 37, 1-34.

Hayes, B., 1995. Metrical stress theory: principles and case studies. The University of Chicago Press, Chicago.

Hayes, B., 1999a. Phonetically-driven phonology: the role of Optimality Theory and Inductive Grounding. In: Darnell, M., Moravcsik, E., Newmeyer, F. J., Noonan, M., Wheatley, K. (Eds.), Functionalism and formalism in linguistics, vol. 1: General papers. John Benjamins, Amsterdam, pp. 243-285. [ROA 158.]

Hayes, B., 1999b. Phonological acquisition in Optimality Theory: the early stages. ROA 327. Revised version published 2004 in: Kager, R., Pater, J., Zonneveld, W. (Eds.), Fixing priorities: constraints in phonological acquisition. Cambridge University Press, Cambridge.

Heck, F., Müller, G., Vogel, R., Fischer, S., Vikner, S., Schmid, T., 2002. On the nature of the input in Optimality Theory. The Linguistic Review 19, 345–376.

Hopper, P.J., Traugott, E.C., 1993. Grammaticalization. Cambridge University Press, Cambridge.

Huang, J., 1984. On the distribution and reference of empty pronouns. Linguistic Inquiry 15, 531–574.

Huang, J., 1989. Pro-drop in Chinese. In: Jaeggli, O., Safir, K. (Eds.), The null subject parameter. Kluwer, Dordrecht, pp. 185–214.

Huang, Y., 1992. Against Chomsky's typology of empty categories. Journal of Pragmatics 17, 1–29.

Huang, Y., 1995a. The syntax and pragmatics of anaphora. Cambridge University Press, Cambridge.

Huang, Y., 1995b. On null subjects and null objects in generative grammar. Linguistics 33, 1081–1123.

Hulst, H. van der, 2000. Modularity and modality in phonology. In: Burton-Roberts, Carr, Docherty (Eds.), pp. 207-243.

Hulst, H. van der, this volume. On the parallel organization of linguistic components.

Hume, E., Johnson, K. (Eds.), 2001. The role of speech perception in phonology. Academic Press, San Diego

Hurford, J.R., 1990. Nativist and functional explanations in language acquisition. In: Roca, I. M. (Ed.), Logical issues in language acquisition. Foris Publications, Dordrecht, pp. 85-136.

Hyman, L.M., 2001. The limits of phonetic determinism in phonology: *NC revisited. In: Hume, Johnson (Eds.), pp. 141-185.

Inkelas, S., Orgun, C.O., 1995. Level ordering and economy in the lexical phonology of Turkish. Language 71, 763-793.

Itô, J., 1989. A prosodic theory of epenthesis. Natural Language and Linguistic Theory 7, 217-260.

Jackendoff, R., 2002. Foundations of language: brain, meaning, grammar, evolution. Cambridge University Press, Cambridge.

Jakobson, R., 1929. Remarques sur l'évolution phonologique du russe comparée à celle des autres langues slaves. Travaux du Cercle Linguistique de Prague 2. Reprinted, 1962, in: Selected writings, vol. I: Phonological studies. Mouton & Co., The Hague, pp. 7-116.

Johnson, K. (1997). Speech perception without speaker normalization. In: Johnson, Mullennix (Eds.), pp. 145-65.

Johnson, K., Mullennix, J.W. (Eds.), 1997. Talker variability in speech processing. Academic Press, San Diego, CA.

Kager, R., 1999. Optimality theory. Cambridge University Press, Cambridge.

Kalmár, I., 1979. Case and context in Inuktitut (Eskimo). National Museum of Man Mercury Series, 49. National Museums of Canada, Ottawa.

Karmiloff-Smith, A., 1998. Development itself is the key to understanding developmental disorders. Trends in Cognitive Sciences 2, 389-98.

Katz, J., Postal, P.M., 1964. An integrated theory of linguistic descriptions. MIT Press, Cambridge, MA.

Keating, P.A., 1984. Aerodynamic modeling at UCLA. UCLA Working Papers in Phonetics 54, 18-28.

Kiparsky, P., 1995. The phonological basis of sound change. In: Goldsmith, J.A. (Ed.), pp. 640-670.

Kirchner, R., 1997. Contrastiveness and faithfulness. Phonology 14, 83-111.

Kuhn, J., 2003. Optimality-theoretic syntax: a declarative approach. CSLI Publications, Stanford, CA.

Ladefoged, P., Maddieson, I., 1996. The sounds of the world's languages. Blackwell, Oxford.

Langacker, R., 1987. Foundations of Cognitive Grammar, vol. 1. Stanford University Press, Stanford, CA.

Legendre, G., 2000. Morphological and prosodic alignment of Bulgarian clitics. In: Dekkers, van der Leeuw, van de Weijer (Eds.), pp. 423-62.
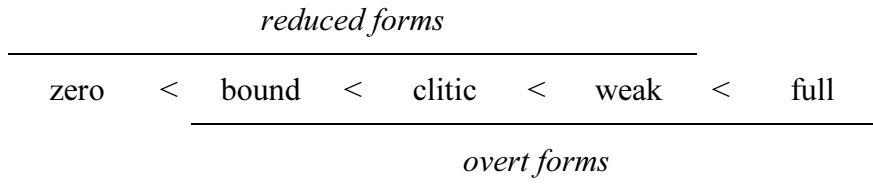
Legendre, G., 2001. Introduction to Optimality Theory in syntax. In: Legendre, Grismshaw, Vikner (Eds.), pp. 1–27.

Legendre, G., Grimshaw, J., Vikner, S. (Eds.), 2001. Optimality-theoretic syntax. MIT Press, Cambridge, MA.

Legendre, G., Smolensky, P., Wilson, C., 1998. When is less more? Faithfulness and minimal links in wh-chains. In: Barbosa, Fox, Hagstrom, McGinnis, Pesetsky (Eds.), pp. 249-89.

Legendre, G., Wilson, C., Smolensky, P., Homer, K., William, R., 1995. Optimality and wh-extraction. In: Beckman, Dickey, Urbanczyk (Eds.), pp. 607-36.

Lightfoot, D., 1999. The development of language: acquisition, change, and evolution. Blackwell, Oxford.

Lindblom, B.E.F., 1990. Explaining phonetic variation: a sketch of the H&H theory. In: Hardcastle, W., Marchal, A. (Eds.), Speech production and speech modeling. Kluwer, Dordrecht, pp. 403-39.

Marcus, G.F., 2001. The algebraic mind: integrating connectionism and cognitive science. The MIT Press, Cambridge, MA.

McCarthy, J.J., 1993. A case of surface constraint violation. Canadian Journal of Linguistics 38, 169-195.

McCarthy, J.J., 2002. A thematic guide to Optimality Theory. Cambridge University Press, Cambridge.

McCarthy, J.J., Prince, A., 1994. The emergence of the unmarked: optimality in prosodic morphology. In: Gonzàlez, M. (Ed.), Proceedings of the North East Linguistic Society 24. Graduate Linguistic Student Association, Amherst, MA, pp. 333-79.

McCarthy, J.J., Prince, A., 1995. Faithfulness and reduplicative identity. In: Beckman, Dickey, Urbanczyk (Eds.), pp. 249-384.

McMahon, A.M.S., 2000. Change, chance, and optimality. Oxford University Press, Oxford.

McMahon, A.M.S., 2003. Phonology and the Holy Grail. Lingua 113, 103-115.

Müller, G., 2001. Order preservation, parallel movement, and the emergence of the unmarked. In: Legendre, Grismshaw, Vikner (Eds.), pp. 279-313.

Myers, S., 1997. Expressing phonetic naturalness in phonology. In: Roca (Ed.), pp. 125-152.

Myers, S., 2000. Boundary disputes: the distinction between phonetic and phonological sound patterns. In: Burton-Roberts, Carr, Docherty (Eds.), pp. 245-72.

Myers, S., 2002. Gaps in factorial typology: the case of voicing in consonant clusters. Ms., University of Texas at Austin. [ROA 509.]

Newmeyer, F.J., 1991. Functional explanation in linguistics and the origins of language. Language and Communication 11, 3-28.

Newmeyer, F.J., 1998. The irrelevance of typology for linguistic theory. Syntaxis 1, 161-197.

Newmeyer, F.J., 2000. Optimality and functionality: some critical remarks on OT syntax. Ms., University of Washington. [ROA 402.]

Newmeyer, F.J., 2002a. Optimality and functionality: a critique of functionally-based optimality-theoretic syntax. Natural Language and Linguistic Theory 20, 43-80.

Newmeyer, F.J., 2002b. A rejoinder to Bresnan and Aissen. Natural Language and Linguistic Theory 20, 97-99.

Nichols, J., 1986. Head-marking and dependent-marking grammar. Language 62, 56-119.

Ohala, J.J., 1983. The origin of sound patterns in vocal tract constraints. In: MacNeilage, P. F. (Ed.), The production of speech. Springer, New York, pp. 189-216.

Ohala, J.J., 1989. Sound change is drawn from a pool of synchronic variation. In: Breivik, L. E., Jahr, E.H. (Eds.), Language change: contributions to the study of its causes. Mouton de Gruyter, Berlin, pp. 173-198.

Ohala, J.J., 1990. There is no interface between phonology and phonetics: a personal view. Journal of Phonetics 18, 153-171.

Ohala, J.J., 1992. What's cognitive, what's not, in sound change. In: Kellermann, G., Morrissey, M. D. (Eds.), Diachrony within synchrony: language history and cognition. Papers from the International Symposium at the University of Duisburg, 26-28 March 1990. Peter Lang, Frankfurt am Main, pp. 309-355.

Ohala, J.J., 1993. The phonetics of sound change. In: Jones, C. (Ed.), Historical linguistics: problems and perspectives. Longman, London/New York, pp. 237-278.

Orgun, C.O., 1996. Sign-Based Morphology and Phonology, with special attention to Optimality Theory. PhD dissertation, University of California, Berkeley.

Orgun, C.O., Sprouse, R. L., 1999. From MPARSE to CONTROL: deriving ungrammaticality. Phonology 16, 191-224.

Pater, J., 1999. Austronesian nasal substitution and other NC effects. In: Kager, R., van der Hulst, H., Zonneveld, W. (Eds.), The prosody-morphology interface. Cambridge University Press, Cambridge, pp. 310-43.

Pesetsky, D., 1997. Optimality Theory and syntax: movement and pronunciation. In: Archangeli, Langendoen (Eds.), pp. 135-170.

Pesetsky, D., 1998. Some optimality principles of sentence pronunciation. In: Barbosa, Fox, Hagstrom, McGinnis, Pesetsky (Eds.), pp. 337-383.

Pierrehumbert, J., 1990. Phonological and phonetic representation. Journal of Phonetics 19, 375-94.

Pierrehumbert, J., Beckman, M.E., Ladd, D.R., 2000. Conceptual foundations of phonology as a laboratory science. In: Burton-Roberts, Carr, Docherty (Eds.), pp. 273-303.

Pinker, S., Bloom, P., 1990. Natural language and natural selection. Behavioral and Brain Sciences 13, 707-84.

Pisoni, D.B., 1997. Some thoughts on 'normalization' in speech perception. In: Johnson, Mullennix (Eds.), pp. 9-32.

Poole, G., 1996. Transformations across components. PhD dissertation, Harvard University.

Poole, G., 1997. Stylistic Fronting in Icelandic: a case study in prosodic X^0-movement. Newcastle/Durham Working Papers in Linguistics 4, 249-285.

Poole, G., 2001. Syntactic vs. phonological displacement: some empirical considerations. Paper presented at the Ninth Manchester Phonology Meeting, Manchester, 25 May 2001.

Prince, A., Smolensky, P., 1993. Optimality Theory: constraint interaction in generative grammar. Ms, Rutgers University Center for Cognitive Science Technical Report #2. Revised (2002), ROA 537.

Reiss, C., 2000. Optimality Theory from a cognitive science perspective. In: Ritter (Ed.), pp. 291-301.

Ringen, C., Gao, B., 2002. On the issue of phonetics in phonology: Chinese diminutive suffixation. Paper presented at the Tenth Manchester Phonology Meeting, Manchester, 23-25 May 2002.

Ritter, N.A. (Ed.), 2000. A review of Optimality Theory. Special issue of The Linguistic Review 17 (2-4), 111-384.

Rizzi, L., 1986. Null objects in Italian and the theory of pro. Linguistic Inquiry 17, 501–558.

Roca, I. (Ed.), 1997. Derivations and constraints in phonology. Clarendon Press: Oxford.

Samek-Lodovici, V., 1996. Constraints on subjects: an optimality-theoretic analysis. Unpublished PhD dissertation, Rutgers University, New Brunswick, NJ.

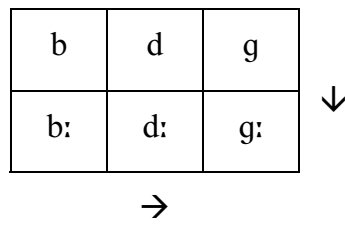Sells, P., 2001a. Structure, alignment and optimality in Swedish. CSLI Publications, Stanford, CA.

Sells, P., 2001b. Form and function in the typology of grammatical voice systems. In: Legendre, Grismshaw, Vikner (Eds.), pp. 355–391.

Speas, M., 1997. Optimality theory and syntax: null pronouns and control. In: Archangeli, Langendoen (Eds.), pp. 171–199.

Speas, M., 2001. Constraints on null pronouns. In: Legendre, Grismshaw, Vikner (Eds.), pp. 393–425.

Steriade, D., 1997. Phonetics in phonology: the case of laryngeal neutralization. Ms., UCLA.

Steriade, D., 2000. Paradigm uniformity and the phonetics-phonology boundary. In: Broe, M. E., Pierrehumbert, J. B. (Eds.), Papers in Laboratory Phonology, vol. V: Acquisition and the lexicon. Cambridge University Press, Cambridge, pp. 313-334.

Steriade, D., 2001. Directional asymmetries in place assimilation: a perceptual account. In: Hume, Johnson (Eds.), pp. 219-50.

Tallerman, M., this volume. Challenging the syllabic model of 'syntax-as-it-is'.

Tesar, B., Smolensky, P., 2000. Learnability in Optimality Theory. The MIT Press, Cambridge, MA.

Thompson, D.W., 1942. On growth and form: a new edition. Cambridge University Press, Cambridge.

Törkenczy, M., 2000. Absolute phonological ungrammaticality: defective paradigms in Hungarian. Handout of paper presented at the 8th Manchester Phonology Meeting, Manchester, 18-20 May 2002.

Uffmann, C., 2002. Intrusive /r/ and optimal epenthetic consonants. Paper presented at the Conference on English Phonology, Toulouse, 26-28 June 2002.

Uriagereka, J., 1998. Rhyme and reason in human language: an introductory dialogue to minimalist syntax. MIT Press, Cambridge, MA.

Vennemann, T., 1972. Rule inversion. Lingua 29, 209-242.

Vennemann, T., 1988. Preference laws for syllable structure and the explanation of sound change, with special reference to German, Germanic, Italian and Latin. Mouton de Gruyter, Berlin.

Vikner, S., 2001. V^0-to-I^0 movement and do insertion in Optimality Theory. In: Legendre, Grismshaw, Vikner (Eds.), pp. 427–464.

Vincent, N., 2000. Competition and correspondence in syntactic change: null arguments in Latin and Romance. In: Pintzuk, S., Tsoulas, G., Warner, A. (Eds.), Diachronic syntax: Models and mechanisms. Oxford University Press, Oxford, pp. 25–50.

Westbury, J., Keating, P., 1986. On the naturalness of stop consonant voicing. Journal of Linguistics 22, 145-166.

Wetzels, W.L., Mascaró, J., 2001. The typology of voicing and devoicing. Language 77, 207-46.

Zhang, J., 2000. Non-contrastive features and categorical patterning in Chinese diminutive suffixation: MAX[F] or IDENT[F]? Phonology 17, 427-478.

DIAGRAMS

    (12)   **Pronominal forms**

reduced forms

zero   <   bound   <   clitic   <   weak   <   full

overt forms

(22) **Performance difficulty landscape for voiced plosives**

b	d	g
bː	dː	gː

↓

→

(23)   **Two possible performance difficulty thresholds among voiced plosives**

a.

b	d	g
bː	dː	gː

b.

b	d	g
bː	dː	gː

(30)   *The law*[l] *is,* but *the bra*[∅] *is*

		*G[-hi]	DEP-VPlace	DEP-[−son]	ONSET	DEP-CPlace	*[lateral]
/lɔː ɪz/	lɔː.ɔ̨ɪz	*!					
	lɔː.wɪz		*!				
	lɔː.ʔɪz			*!			
	lɔː.ɪz				*!		
	lɔː.lɪz ☞					*	*
/bɹɑː ɪz/	bɹɑː.ɑ̨ɪz	*!					
	bɹɑː.wɪz		*!				
	bɹɑː.ʔɪz			*!			
	bɹɑː.ɪz ☞				*		
	bɹɑː.lɪz		*!			*	*

NOTES

[1] On García (1979) and Kalmár (1979) see Croft (1995: §6). García denies the existence of arbitrariness in language beyond the Saussurean relationship between sound and meaning in individual words. No view like this has been articulated in more recent research, but García's position is still often brought forth as a straw man.

[2] These programmatic assumptions have increasingly led Minimalists to reduce the scope of syntax by relegating to the PF component many phenomena traditionally regarded as syntactic. We discuss this trend below in connection with the proposals made by Burton-Roberts & Poole (this volume).

[3] According to Burton-Roberts & Poole, many phenomena traditionally regarded as syntactic are in fact phonological (see the previous footnote). They do not however deny that these supposedly phonological phenomena are different in nature from the facts studied by traditional or 'narrow' phonology. Their arguments against Poole's (1996, 1997, 2001) analysis of Stylistic Fronting in Icelandic presuppose such a distinction.

[4] In fact, we do not at present know exactly what the innate language faculty contains under the Representational Hypothesis. For example, Burton-Roberts & Poole (this volume: §6.1) retain V-to-I movement in the 'narrow syntax', if only as a provisional expedient. In our view, however, the internal logic of the Representational Hypothesis leaves no room for finiteness and V-to-I raising in the language faculty. Owing to this sort of uncertainty, it is not yet possible to subject the Representational Hypothesis to comprehensive empirical testing.

[5] See also Bermúdez-Otero (2002: 409). Similarly, Belletti and Rizzi wonder whether the pursuit of Minimalist assumptions may conflict with the goal of solving the logical problem of language acquisition (Chomsky et al., 2002: 132).

[6] In §6.3 we shall argue in favour of the received view that constraints refer only to specifically phonological categories. In §7, in contrast, we suggest that it is implausible to claim that all members of CON are supplied innately by UG, and we survey scenarios where markedness constraints are ontogenetically emergent.

[7] In opposition to CODACOND-Place, Steriade (1997) suggests that the markedness constraints driving place neutralization should directly refer to the phonetic phenomena (such as acoustic transitions) that cue place contrasts. We criticize this proposal in §6.3.

[8] Notably, Heck et al. (2002) suggest that in a transformational framework it is possible to dispense with syntactic input representations —and, concomitantly, with syntactic faithfulness constraints— because traces and abstract features provide a record of syntactic derivations that is directly accessible to output constraints (though see §5.1). In LFG-OT, in

contrast, these representational devices are not available; rather, the basic organization of the theory in terms of f-structure and c-structure representations entails a genuine input-output relationship. In any case, this issue is not crucial to the more basic question whether, like phonology, syntax is governed by markedness constraints; see below.

[9] Parsing is an area where the issues do not appear intractable: see in particular the work of Hawkins (1989, 2001).

[10] There are parallels to this debate in phonology: McMahon (2003) suggests that optimality-theoretic constraints are not appropriate to all phonological phenomena; she argues that only prosody (particularly intonation) is of a nature suitable for analysis in OT terms.

[11] It should be pointed out here that Pesetsky (1997) repeatedly warns that his argumentation and conclusions are tentative and controversial.

[12] There are remarkable similarities between Pesetsky's proposals and those of Burton-Roberts & Poole (this volume). Burton-Roberts & Poole divide the province of traditional syntax between the generation of syntactic objects proper and the conventional m-representation of those objects in the phonetic medium. Significantly, they assume that syntactic objects contain chains and that m-representational conventions determine how the members of a chain are phonetically m-represented. These proposals are reminiscent of Pesetsky's distinction between movement and pronunciation. In addition, some of the m-representational conventions that Burton-Roberts & Poole formulate in their analysis of Icelandic Stylistic Fronting are in conflict with each other, in that they cannot all be satisfied simultaneously; in this situation, Burton-Roberts & Poole identify certain conventions as taking precedence. This is close to Pesetsky's claim that 'pronunciation' is controlled by ranked optimality-theoretic constraints. Similar approaches to the relationship between chains and pronunciation are pursued in Minimalist works such as Brody (1995) and Groat & O'Neil (1996); these authors, however, do not exclude phonology from UG.

[13] There are some problematic gaps in the $*NC_{[-voi]}$ conspiracy, for, in addition to (2), factorial typology predicts a few repair strategies that appear not to be attested (Pater, 1999; Myers, 2002). As we shall see in §6.1, this problem has a direct bearing on the general question concerning the nature and origin of markedness constraints.

[14] This example presupposes a parallel-correspondence approach to grammatical relations (see §3). The variety of strategies for subject marking shown in (4b) is obscured in theoretical frameworks which, like GB, assume structural definitions of grammatical relations for all languages; but we do not find this assumption warranted. In the following paragraphs, in any case, we provide examples of syntactic conspiracies which either arise with equal clarity in all theoretical frameworks or specifically presuppose a transformational perspective.

[15] In GB or Minimalism the subject would of course be marked with a nonrealized nominative Case. This does not hold for the approach to syntactic representations that we assume here.

[16] The subordinate constraint ALIGN(Left, V_{fin}; Left, Clause) pushes the finite verb as far left in the clause as possible, but top-ranked NONINITIAL(V_{fin}) prevents it from occurring in absolute clause-initial position. We use the term 'clause' here to avoid controversial analytical commitments. The commonest assumption would be that what we term 'clause' would correspond to a CP. However, we do not wish to exclude the possibility that, in V2 languages, clauses may be flat, perhaps non-headed, structures —say S. Under the analysis proposed by S. Anderson (1996, 2000), for example, some of the empirical arguments in favour of C as the head of these clauses are weakened. Subordinate clauses do not display V2 effects in German; we assume that this is due to other overriding constraints.

[17] For similar facts in Icelandic, see Burton-Roberts & Poole (this volume).

[18] What counts as a zero pronoun varies between frameworks. In the Chomskyan tradition, the term 'zero pronoun' refers to the absence of an overt syntactic pronoun. In LFG, in contrast, a distinction is made between bound pronouns and agreement; constructions involving bound pronouns are not considered to be instances of pro-drop (Bresnan, 2001a: ch 8).

[19] Though see Gilligan (1987) for problems with the link between pro-drop and agreement.

[20] See also Burzio (1998), who argues that his analysis of the behaviour of anaphoric elements in terms of hierarchies can be fruitfully translated into soft constraints.

[21] To be absolutely precise, epenthetic segments vacuously satisfy IDENT-feature and MAX-feature constraints. However, they do violate DEP-feature constraints, as their featural content is not present in the input. See McCarthy & Prince (1995) for the formulation of faithfulness constraints. In §6.5 we analyse a case of segmental epenthesis in which DEP-feature constraints play a crucial controlling rôle; see tableau (30).

[22] The intrusive *r* of several nonrhotic English dialects has been claimed to constitute a counterexample to the emergence of the unmarked. We address this controversial case below in §5.3 and §6.5.

[23] As we pointed out in §3, Heck *et al.* (2002) claim that inputs and faithfulness constraints need not play a rôle in OT syntax, at least within implementations allied to GB or Minimalism; by implication, they adhere to the Katz-Postal Hypothesis for GEN. We note, however, that Heck *et al.* do not address the arguments related to ineffability.

[24] There are several well-rehearsed arguments against stochastic evaluation (see e.g. Newmeyer, 2002a, b), but many of them can be easily countered. Thus, insofar as a stochastic grammar embodies an individual speaker's knowledge of variation, stochastic evaluation in no way commits one to accepting the existence of 'social grammars'. Similarly, the familiar move whereby statistical patterns of variation are dismissed as mere performance effects raises problems in cases where the extent and limits of variation are language-specific.

[25] For a recent discussion of phonologization from the viewpoint of OT, see Bermúdez-Otero & Hogg (2003: §1.1).

[26] Our discussion of rule inversion here and in §6.5 is closely based on Bermúdez-Otero & Hogg (2003: §1.2), to which we refer the reader for further explanation and references to the literature.

[27] CV syllables are a case in point: there is no hierarchization of CON under which CV syllables are not possible in at least some environment. On the apparent counterevidence from Arrernte, see Evans (1995: §5.1) and references therein.

[28] We assume that, in the learner's initial state, all markedness constraints outrank all faithfulness constraints (for references, see McCarthy, 2002: 231; for criticism of this assumption, see Hale & Reiss, 1998). In §7.2 we discuss an approach to the origin of markedness constraints where this initial ranking holds without stipulation.

[29] We borrow this term from Hurford (1990: §2.1).

[30] The argument developed in this section relies heavily on Hayes (1999a: §6.2).

[31] For criticism of Zhang (2000), see Ringen & Gao (2002).

[32] In the sense of SPE (Chomsky & Halle, 1968).

[33] Our exposition in §6.4 and §6.5 develops ideas advanced in Bermúdez-Otero & Hogg (2003: §1.1, §1.2).

[34] See for example Beddor, Krakow & Lindemann's (2001) discussion of listeners' partial compensation for coarticulatory effects.

[35] As we saw in §6.1, VOP prohibits voiced obstruents. *[-son, +con] penalizes continuant obstruents, i.e. fricatives. In turn, the faithfulness constraint IDENT-voice requires each

segment to bear the same voice specifications as its input correspondent, and therefore preserves voice contrasts.

[36] Our arguments here expand and clarify points made in Bermúdez-Otero & Hogg (2003: §1.2), to which we refer the reader for further discussion.

[37] For the restructuring of input representations in phonology, see Bermúdez-Otero & Hogg (2003: §2).

[38] The standard of comparison here is the general faithfulness constraint DEP-VPlace. From the viewpoint of the more specific constraint DEP-[±high], it would be better to insert [l] after /ɑː/ or /ə/ than to epenthesize [j] or [w], for [l] can get its [−high] specification by spreading from the preceding vowel. Thus, the ranking { *G$_{[-hi]}$, DEP-[±high], DEP-[−son], ONSET » DEP-VPlace, DEP-CPlace, *[lateral] } predicts that *l*-intrusion is indeed possible after /ɑː/ or /ə/, on the grounds of partial homorganicity. This prediction is borne out in a minority of *l*-intruding dialects (Gick, 1999: 37). By factorial typology, however, *l*-intrusion is predicted to be impossible after high vowels, at least in languages that tolerate high glides. Again, the facts of *l*-intrusion in English support this conclusion.

[39] Carr (this volume) explicitly assumes that the syllable is not an innate category but rather emerges through a developmental process in which babbling plays a crucial rôle.

[40] For clarification of the epistemological background, see Karmiloff-Smith's (1998) admirably perspicuous comparison of nativist, empiricist, and neuroconstructivist approaches to cognitive psychology.

[41] In the following discussion Boersma's (1998b) suggestions concerning the origin of phonological markedness constraints are adapted to the overall theoretical framework that we have developed in previous sections. Certain other aspects of Boersma's theory of 'Functional Phonology' are incompatible with our proposals: notably, Boersma largely rejects the existence of specifically phonological categories (see Boersma, 1998a).

[42] Hayes (1999b: footnote 15) endorses this view with respect to GEN, which he describes as 'the most idealized aspect of OT'.

[43] The innate specifications countenanced by neuroconstructivists include brain macrostructures (but not neural microcircuitry), attention biases, and domain-relevant (though not necessarily domain-specific) predispositions.