



An invitation to OT syntax and semantics

Vieri Samek-Lodovici

*University College London, Italian Department, Gower Street,
London WC1E 6BT, UK*

Received 10 June 2006; received in revised form 10 June 2006; accepted 16 June 2006
Available online 16 January 2007

Abstract

This brief introduction to the volume compares models of grammaticality based on the simultaneous satisfaction of all UG constraints (e.g. Principles and Parameters, Minimalism) with models based on constraint conflict (e.g. Optimality Theory). It examines the consequences that these alternative definitions have on the analysis of crosslinguistic variation, economy of movement, and the conflict between economy of movement and economy of structure. In particular it shows how the desire to keep a simple definition of grammaticality (i.e. one based on simultaneous constraint satisfaction) is paid by the hidden complexity of economy principles and the theory-internal split separating the theoretical components addressing variation from those addressing linguistic universals, making variation accidental. In contrast, defining grammaticality on the base of constraint conflict roots variation into UG while keeping constraint complexity at check. A brief survey of the articles collected in this volume completes this introduction.

© 2006 Elsevier B.V. All rights reserved.

Keywords: Optimality Theory; Minimalism; Grammaticality; UG; Economy; Crosslinguistic variation

1. Introduction

A definition of grammaticality is an essential prerequisite of generative studies. Without it we cannot extract knowledge about the content of UG-constraints from our empirical observations. Yet this definition depends on initial assumptions about the UG constraints themselves. Is the content of UG-constraints such that they never conflict with each others, making it possible to simultaneously satisfy them all? Or is their content such that at least some constraints conflict with one another, making it impossible to satisfy all of them at once?

The answer is obviously an empirical issue. We do not know a priori the content of UG constraints, and hence we also do not know a priori whether they conflict or not. Yet the answer we choose determines our notion of grammaticality. If we assume that constraints do not conflict, then grammaticality can be defined as the simultaneous satisfaction of all constraints, as in the Principles and Parameters and Minimalist frameworks (Chomsky, 1981, 1995). If we assume that constraints conflict, simultaneous satisfaction becomes untenable and grammaticality must follow a definition that allows for constraint violation such as the one proposed in OT (Prince and Smolensky, 1993/2004).¹

It is instructive to examine how this initial assumption on constraint conflict affects the formal devices with which distinct UG-models capture empirical observations. I will here argue that the existence of crosslinguistic variation and the existence of robust economy-centered empirical generalizations are captured naturally and with no further assumptions or formal devices within a conflict-based approach like OT, whereas they force complex and unnecessary formal assumptions under simultaneous constraint satisfaction approaches.

2. The impact of grammaticality definitions on UG-models

There are at least three important aspects of UG that are immediately affected by our view of UG-constraints and the related definitions of grammaticality. They concern the analysis of crosslinguistic variation, the complexity of constraint definitions, and the modeling of constraint interaction. Each aspect is discussed below in turn.

2.1. Crosslinguistic variation

There is an inevitable tension underlying generative studies of human language. On one hand we want to identify and characterize the universal constraints at the core of UG, yet our model must also account for the crosslinguistic variation affecting almost every aspect of grammar. The solution to this tension directly depends on our initial assumption on the conflicting nature of UG-constraints and provides very different answers to the questions of why crosslinguistic variation exists and how it is related to UG.

Under the OT model of constraint conflict, crosslinguistic variation is rooted in and determined by UG. Crosslinguistic variation constitutes the expected necessary consequence of the different possible resolutions of the conflicts among UG constraints (Prince and Smolensky, 1993/2000; Grimshaw, 1997:405). Each distinct resolution yields a crosslinguistic variant. It follows that in this model UG constraints directly determine *whether*, *where*, and *how* variation occurs. They determine *whether* variation occurs, because variation may only occur whenever

¹ In OT distinct languages correspond to distinct rankings of UG constraints and linguistic structures compete freely against each other. Intuitively, a structure is optimal and hence grammatical when it maximizes constraint satisfaction relative to a specific ranking by keeping constraint violations at the absolute minimum necessary to solve any available constraint conflict. Which constraint is violated is determined by the given ranking. Slightly more formally, a structure is optimal in a language when it remains unbeaten by any other competitor when assessed with respect to the ranking of that language. A structure A beats another structure B whenever A outperforms B on the highest constraint on which A and B differ (Prince and Smolensky, 1993/2004; Grimshaw, 2005).

two or more constraints conflict with each other.² They determine *where* it occurs because variation may and must occur only with respect to those aspects of grammar on which some constraints conflict; where constraints do not conflict no variation ensues. They determine *how* variation occurs because the properties of the structures providing the possible resolutions of each specific conflict are those dictated by the conflicting constraints themselves; nothing else can determine these structural properties.

The same cannot be said for models based on simultaneous constraint satisfaction, where crosslinguistic variation appears to be severed from UG constraints. Left on their own, non-conflicting UG constraints necessarily predict crosslinguistic convergence on perfect grammatical forms that satisfy them all, the exact opposite of crosslinguistic variation. Determining *whether*, *where*, and *how* variation occurs then inevitably falls on the shoulders of additional and external parametric devices such as P&P's parameters or Minimalism's feature properties (whether parametric strength, or differences in feature distribution across syntactic structures, lexemes, or languages). The overall model of UG is thus effectively split into two distinct components, one accounting for universal properties (UG constraints) and one for crosslinguistic variation (parametric devices). Unlike the constraint-conflict model discussed above, no logical necessity links the parametric component to the universal constraints of grammar, making the occurrence of variation surprising, an accidental property external and unrelated to the universal core of UG.

2.2. Constraint complexity

In examining the consequences brought by our initial assumption on the conflicting nature of UG constraints, we have to ensure that we compare like with like. This can be problematic because a constraint that appears to be obviously conflicting under a constraint conflict approach like OT can be recast as an inviolable one under a more complex definition that allows for all and only the conflict-determined violations imposed by more prominent constraints (Grimshaw, 1997:405; Burzio, 2000).

This is typically the case with economy principles. As Cardinaletti and Starke (1999:210) explicitly point out, minimalist principles of the form '*Minimise α* ' should be understood as '*Minimise α up to crash*', thus implying that any failure to minimize α necessary to avoid crashing should not be counted as a violation, while any additional failures will. The same is true for other principles requiring comparisons among possible structures or derivations. Consider for example the principle *Procrastinate* which disfavors overt movement in early Minimalism

² Strictly speaking variation is also contingent on the existence of distinct rankings for the conflicting constraints. If the ranking of the constraints is universally fixed no variation ensues. The null hypothesis, however, is that no ranking is superior to any other, thus supporting free ranking of UG constraints and the entailed crosslinguistic variation.

Fixed rankings do occur in plenty of OT-analyses, including those in this volume. These fixed constraint hierarchies however are determined in a systematic, non-arbitrary, way from equally fixed markedness scales concerning specific classes of grammatical features or functional items. These markedness-scales are genuine empirical discoveries that must be retained by any model of UG. Far from calling into question the rooting of crosslinguistic variation within UG, fixed hierarchies confirm it by predicting typologies of specific kinds when conflicting with faithfulness constraints. In particular, the lower the rank of faithfulness constraint within a fixed hierarchy of markedness constraints, the less marked the structures of the corresponding language will have to be, because faithfulness to input features will not be allowed to override any of the higher ranked markedness constraints. This gives rise testable markedness implicatures where the existence of a marked structure or item in a language entails the presence of all the other less marked structures or items within the same language.

(Chomsky, 1995:198). The principle plays an important role in blocking overt movement in languages where the latter is not attested. For example, the alternation between V-to-T in French and its absence in English is modeled by positing a strong feature which requires overt checking in T^o despite Procrastinate on French verbs, and a weaker one whose checking must instead obey Procrastinate on English verbs.

The violable status of Procrastinate depends on its definition. If we define it simply as ‘*no overt movement*’, then Procrastinate is violated whenever feature checking forces overt movement, whether the relevant feature is weak or strong. This definition reveals the underlying conflict between Feature-Checking of strong features and Procrastinate, with Procrastinate violated whenever Feature-Checking requires it. If we instead define Procrastination as ‘*LF-movement is “cheaper” than overt movement*’ (Chomsky, 1995:198, double quotes as in the original) then Procrastinate may be interpreted as an inviolable principle where the violations imposed by feature checking of strong features are ignored, and violations are restricted to overt movement of weak features.

Does assuming or denying constraint-conflict then concern nothing more than the choice between formally distinct but intuitively equivalent definitions of a constraint? No, because the two definitions are not as equivalent as they appear at first sight. This emerges most clearly when considering their assessment. The first definition, ‘*no overt movement*’, can be directly assessed on any given structure regardless of the existence of other constraints: if overt movement occurs, whatever its cause, the constraint is violated.

Chomsky’s definition instead must assess which among the available overt movement operations are dictated by the checking of strong features, and as such assumed to not violate the constraint, and which are not so dictated and hence truly violate the constraint. The ranking between Feature-Checking and Procrastinate is thus an essential component of the assessment of Procrastinate. The conflict-resolution process formalized constraint-externally in OT becomes part of the assessment of the constraint.

Assuming constraint conflict thus enables us to keep constraints genuinely simple while at the same time explaining economy effects in terms of constraint interaction (Grimshaw, 1997, 2005; Prince, 1997:2; Burzio, 2000:209,216; McCarthy, 2002:40; Smolensky et al., 2006:505,531). Contrary to common belief conflict-based models like OT thus favor a genuine and substantial bound on constraint complexity. In contrast, inviolable constraints enforcing economy of some derivational aspect or structural property *state* but do not *explain* the observed economy and pay for their inviolable status with the inherent complexity of their assessment.

2.3. Conflict among economy constraints

The set of theoretical assumptions necessary to keep economy principles formally inviolable becomes even richer and more articulated when economy principles conflict with each other, revealing their fragile status as a means to uphold simultaneous constraint satisfaction while capturing economy-centered empirical generalizations.

A case in point concerns the interaction between structural and movement economy discussed in Cardinaletti and Starke (1994, 1999). On the basis of the different distributions associated with pronouns of different structural complexity, Cardinaletti and Starke identify a robust generalization according to which pronouns with the *least* structure move the *most* in order to visit those functional projections in the clause spine that can provide them with the functional feature-values that they lack. In contrast, the *most* structured pronouns move the *least* because they already encode the relevant feature values in functional projections of their own and hence

need not seek them elsewhere. Most interestingly, these authors also observe a strict preference for pronouns with less structure over more complex ones whenever both pronouns are potentially possible. This leads them to posit a principle of ‘Economy of Representations’ defined as ‘*minimise structure*’ (Cardinaletti and Starke, 1994:38, 1999:198). As they point out, structural economy and movement economy appear to conflict with one another: less structure forces more movement and vice versa, with economy on one principle involving diseconomy on the other. The challenge is how to model this conflict in either approach while at the same time deriving Cardinaletti and Starke’s main generalization.

From a conflict-based perspective their generalization follows straightforwardly from a tripartite ranking where the constraints responsible for feature checking dominate a lower constraint penalizing structure in pronominal forms, itself dominating a constraint penalizing movement such as Stay (Grimshaw, 1997). The highest feature-checking constraints block all derivations crashing on unchecked features, while the rest of the ranking ensures that movement minimization is subordinated to structural minimization of pronominal forms. When two pronominal forms are possible, both satisfying feature checking, the constraint against pronominal structure selects the least structured pronoun despite the additional violations of Stay caused by its longer chain.

Cardinaletti and Starke (1999:202) recognize the potential conflict between the two economy constraints but dispute that an explicit ranking – and the associated violation of movement economy – is necessary under a simultaneous constraint satisfaction approach such as Minimalism. According to them the two principles apply at different stages of the derivation for independent reasons, with economy of representations (ER) applied at lexical insertion and economy of movement (EM) applied as the derivation unfolds and crucially after lexical insertion because no chains occur at the point of lexical insertion. This intuitively suggests a two-step derivation with ER selecting the least structured pronoun P at step 1, followed by EM selecting the derivation with shortest chains among those involving P at step 2. The conflict is avoided because EM always examines derivations sharing the same pronoun, hence no ranking is necessary.

The real complexity of this apparently simple proposal only emerges when considering the assessment of the involved economy constraints. A closer inspection shows that it is not possible to assess ER *prior* and *independently* of EM. At the point of lexical insertion the choice between a more complex strong pronominal form S and a less structured weak form W is contingent on the existence of a non-crashing derivation for W: in other words ER selects W over S only if W’s derivation does not crash. Assessing ER thus inherently requires the unfolding of W’s derivation, including the assessment of EM. It follows that there cannot exist a point of lexical insertion that genuinely precedes the full unfolding of syntactic derivations. Furthermore, step 2 emerges as a redundant and unnecessary repetition of the derivation’s unfolding, since as we just saw its unfolding is a prerequisite of the assessment of ER at lexical insertion.

Most importantly, it becomes apparent that serialization is not a viable strategy for avoiding the conflict between the two economy principles: they must apply at the same stage of the derivation for the reasons just discussed. The only way to solve their conflict while avoiding an explicit ranking of the constraints requires an explicit model of the assessment of ER. This model must specify that ER checks the derivations of the different competing pronominal forms for potential crash independently of one another and crucially always ensuring that each application of EM is restricted to derivations with the same pronoun. The ranking and the associated violability of EM is thus avoidable only by explicitly modeling the subordination of EM to ER in the procedure responsible for ER’s assessment.

As in the previous cases we see unnecessarily complex assumptions – here introducing a highly articulated procedure for the assessment of a specific economy constraint – that follow only from the desire to maintain the view that UG constraints are inviolable. The same conflict and its effects are instead straightforwardly accounted for under a constraint-conflict approach to grammar, with no need for additional assumptions on how the assessment of specific constraints is to be carried out.

3. This volume

The above discussion provides some general and I believe important reasons supporting an OT-approach to grammar (see also Samek-Lodovici, *in press*). Allowing for constraint conflict permits us to root crosslinguistic variation firmly into UG while at the same time deriving the pervasive economy relations uncovered in the last two decades of research from simple constraints whose assessment remains straightforward.

This volume brings together some of the latest research in OT-syntax and syntax–semantics interface. Most of the collected studies have been first presented at the LSA workshop on ‘OT-Syntax and Semantics’ organized by Geraldine Legendre at Harvard’s Linguistic Institute in July 2005. All essays well illustrate the explanatory power of OT, deriving highly articulated and some times truly surprising paradigms (at least for this reader) from simple sets of universal constraints that also account for more familiar patterns once their ranking is changed.

The essays have been ordered according to their topic, starting with Legendre’s analysis of auxiliary selection, followed by Badecker’s and Reilly’s essays on agreement, followed by essays on case-assignment by Woolford, and Yang and van Bergen, and finally ending with two essays on the syntax–semantics interface by de Hoop and Malchukov, and Farkas and de Swart. I am particularly grateful to all the above contributors and the extremely helpful anonymous referees that reviewed their papers for making this volume possible. A brief summary of each essay follows below.

Geraldine Legendre—Legendre’s study examines the notoriously complex crosslinguistic distribution of auxiliaries *have* and *be* across two dimensions: (i) variation in the class of monadic verbs associated with either auxiliary under simple active compound tenses, and (ii) variation in the selected auxiliary across distinct syntactic contexts (such as passives, intransitives, transitives, etc). Following Sorace (2000), monadic verbs are classified according to the set of lexical-semantic features responsible for their aspectual properties. Legendre successfully derives auxiliary choice from the conflict between a first set of constraints governing the tagging of arguments as external or internal and two additional constraints, one favoring *have* for external arguments the other discouraging *have* as the marked auxiliary. By exploiting constraint conflict Legendre succeeds in modeling auxiliary choice across a great variety of languages and syntactic contexts in terms of a simple set of universal constraints and no appeal to language specific rules.

William Badecker—Badecker’s analysis examines partial agreement (agreement with just one of two conjuncts) and a cluster of properties associated with it. By distinguishing between concord agreement (agreement with morphologically specified features) and index agreement (agreement with semantic/interpretative features), Badecker successfully links different kinds of partial agreement and their optional vs. obligatory status in specific contexts to the interaction of constraints on local and extended agreement independently necessary for the analysis of agreement impoverishment under non spec-head configurations. Besides its cross-linguistic

import, the article well illustrates the relevance of harmonic bounding in determining empirical generalization for specific classes of languages, using two simple alignment constraints to derive partial agreement with the NP-conjunct closest to the agreeing head.

Ehren Reilly—Reilly examines highly complex agreement paradigms including languages that encode agreement via the simultaneous use of agreement clitics and verbal agreement affixes. The distribution of agreement affixes and clitics in some of these paradigms resembles that of case in ergative case systems, suggesting a possible link between agreement and ergative case for these occurrences. After examining the evidence against this connection, Reilly proposes a new analysis where the examined distributions emerge from a three-fold tension between constraints requiring overt agreement with an argument's φ -features, markedness constraints governing the way agreement is encoded (whether via clitics or agreement affixes), and finally constraints governing morphological realization, crucially involving adjacency between verbal stems and agreement morphemes (whether clitic or affixes). For example, under the appropriate ranking the impossibility of aligning two clitics at once with the same stem yields agreement systems with ACC-clitics and NOM-related agreement affixes in transitive verbs, but also NOM-clitics in intransitive verbs. As Reilly argues, the proposed interaction between morphology and syntax allows him to derive highly complex agreement distributions in terms of simple constraints capturing universal aspects of morphological and syntactic representations.

Ellen Woolford—Woolford's study examines relativized minimality effects with respect to case-licensing. Crucially, case-licensing failure is shown to occur not only when an independent target occurs closer to a case-licensing head than the intended target but also when an independent target is simply *sufficiently* close to the licensing head even though further away than the intended target. Pursuing a revealing analogy with anaphoric binding, case-licensing is shown to rely on a syntactic domain (namely the smallest domain containing a licenser and a licensee), plus the further condition that this domain be pure, i.e. containing no additional case-assignees. As Woolford shows, under the appropriate syntactic configurations this pure-domain constraint conflicts with well-established universal constraints of case-markedness, determining cross-linguistic alternations concerning the absence/presence of locality effects in case-licensing with no appeal to language specific case-licensing rules. The study ends with a set of interesting remarks about the fruitful insights brought by the OT-analysis on the proper modeling of relativized minimality effects and the similarities between case- and anaphoric-licensing.

Ning Yang and Geertje van Bergen—Yang and van Bergen examine an unusual case of differential object marking (DOM) in Mandarin Chinese. Objects can and in some contexts must scramble preverbally, where they are optionally or obligatorily preceded by a case marking particle according to their degree of animacy and definiteness. What is particularly striking is the apparent reversal of the role played by familiar markedness scales: in contrast with Aissen (2003) where the least definite objects are the most likely to avoid overt case marking, Mandarin Chinese marks them obligatorily while leaving marking optional for more definite objects. Yang and van Bergen show how the same markedness scales and harmonic alignment technique used by Aissen in her seminal crosslinguistic analysis of DOM can also account for the surprising Mandarin Chinese paradigm, provided that an additional scale encoding the higher prominence of preverbal positions relative to postverbal ones is formally introduced. The interesting underlying intuition is that preverbal positions are marked ones for objects and unmarked ones for subjects, and consequently definiteness becomes the unmarked property of items in preverbal positions (as indeed it is for subjects) and indefiniteness the marked property, causing overt marking.

Helen de Hoop and Andrej Malchukov—De Hoop and Malchukov examine a great variety of case paradigms where the association of specific cases to semantic/pragmatic content (namely agentivity/volition) can be distinguished from the use of the same cases for grammatical function (such as the marking of subjects vs. objects). The examined patterns are particularly interesting because language after language the constraint ranking that is necessary to determine the correct case paradigm with respect to grammatical function appears systematically challenged by the semantic use of the same case system, which would seem to favor the opposite ranking. As de Hoop and Malchukov note, a solution within unidirectional optimization is always possible, but inevitably involves positing constraints specific to the subclass of lexemes that show an exceptional behavior case-wise. De Hoop and Malchukov show how in these cases bidirectional optimization provides a far more desirable solution, where the limited but systematic exceptions to the default pattern follow from the same constraints and ranking responsible for the default pattern itself. As is typical of OT analyses, this study thus reduces instances of cross-linguistic as well as language-internal variation to the conflict of a limited set of universal constraints. At the same time it raises an important question over syntactic optimization, showing how bidirectional optimization might be superior to mono-directional optimization in deriving highly articulated patterns from very few and very general constraints.

Donka Farkas and Henriëtte de Swart—Farkas and de Swart's contribution explores the role played by maximality (+max) and familiarity (+fam) features in determining the syntactic expression of plural generics. They identify two universal constraints, one favoring definite articles for expressing [+max] nominals and the other disfavoring them for [–fam] nominals. Their ranking determines the crosslinguistic divergence found in the expression of [+max, –fam] nominals, with languages like English and Dutch choosing bare plurals over definite articles and languages like French and Hungarian choosing definite articles over indefinite forms. Farkas and de Swart also show how the universal nature of these constraints gives rise to emergence of the unmarked effects in both sets of languages, with bare plurals becoming acceptable in French and Hungarian for [–max, –fam] generic nominals and definite articles becoming grammatical in English and Dutch for [+max, +fam] generic nominals. It is this kind of effects that defies parametric analyses, because parametric values as simple as the proposed constraints inevitably fail to derive the emergence of the unmarked effects.

References

- Aissen, J., 2003. Differential object marking: iconicity vs. economy. *Natural Language and Linguistics Theory* 21, 435–483.
- Burzio, L., 2000. The rise of optimality theory. In: Cheng, L., Sybesma, R. (Eds.), *The First Glot International State-of-the-Article Book*. Mouton de Gruyter, New York, pp. 199–220.
- Cardinaletti, A., Starke, M., 1994. *The Typology of Structural Deficiency. A Case Study of the Three Classes of Pronouns*. Ms. MIT, Boston.
- Cardinaletti, A., Starke, M., 1999. The typology of structural deficiency. A case study of the three classes of pronouns. In: van Riemsdijk, H. (Ed.), *Clitics in the Languages of Europe*. Mouton de Gruyter, New York, pp. 145–234.
- Chomsky, N., 1981. *Lectures on Government and Binding*. Foris, Dordrecht.
- Chomsky, N., 1995. *The Minimalist Program*. MIT Press, Cambridge, MA.
- Grimshaw, J., 1997. Projections, heads, and optimality. *Linguistic Inquiry* 28, 373–422.
- Grimshaw, J., 2005. In: *How OT Works: Questions and Answers, Descriptions and Explanations*. Talk handout, DEAL Conference, Zentrum für Allgemeine Sprachwissenschaft, Berlin.
- McCarthy, J., 2002. *A Thematic Guide to Optimality Theory*. Cambridge University Press, Cambridge.
- Prince, A., 1997. *Endogenous Constraints in Optimality Theory*. Talk handout, LSA-Institute, Cornell. Downloadable at <http://ling.rutgers.edu/gamma/talks/instdt1.pdf>.

- Prince, A., Smolensky, P., 1993/2004. *Optimality Theory: Constraint Interaction in Generative Grammar*. Blackwell, Malden, MA/Oxford, UK/Carlton, Aus.
- Samek-Lodovici, V., in press. *Optimality Theory and the Minimalist Program*. In: Vogel, R., Broekhuis, H. (Eds.), *Optimality Theory and Minimalism: a Possible Convergence? Linguistics in Potsdam*, University of Postdam.
- Smolensky, P., Legendre, G., Tesar, B., 2006. *Optimality theory: the structure, use, and acquisition of grammatical knowledge*. In: Smolensky, P., Legendre, G. (Eds.), *The Harmonic Mind*, vol. 1. MIT Press, Cambridge, MA, pp. 453–544.
- Sorace, A., 2000. Gradients in auxiliary selection with intransitive verbs. *Language* 76, 859–890.