Constraints and Resources in Natural Language Syntax and Semantics

Gosse Bouma, Erhard W. Hinrichs,
Geert-Jan M. Kruijff, and Richard T. Oehrle

July 23, 1999

CENTER FOR THE STUDY
OF LANGUAGE
AND INFORMATION

Contents

Introduction

Gosse Bouma, Erhard Hinrichs,
Geert-Jan Kruijff, and Richard Oehrle

The papers in this volume represent a broad spectrum of contemporary theoretical and empirical linguistic concerns. Although they flow from somewhat different research paradigms and draw on a variety of different perspectives, they reflect the fact that the paradigms and perspectives represented share important common ground, especially with regard to the role of logical, mathematical, and computational concepts and techniques in linguistic analysis.

A touchstone of these views can be found in the famous passage from the preface of Chomsky (1957), which emphasizes the heuristic role of formalization in clarifying linguistic analyses, especially in a way that will support empirical testing and possible falsification.

> ... The search for rigorous formulation in linguistics has a much more serious motivation than mere concern for logical niceties or the desire to purify well-established methods of linguistic analysis. Precisely constructed models for linguistic structure can play an important role, both negative and positive, in the process of discovery itself. By pushing a precise but inadequate formulation to an unacceptable conclusion, we can often expose the exact source of this inadequacy and, consequently, gain a deeper understanding of the linguistic data. More positively, a formalized theory may automatically provide solutions for many problems other than those for which it was explicitly designed. Obscure and intuition-bound notions can neither lead to absurd conclusions nor provide new and correct ones, and hence they fail to be useful in two important respects.

Constraints and Resources in Natural Language Syntax and Semantics.
Gosse Bouma, Erhard W. Hinrichs, Geert-Jan M. Kruijff, and Richard T. Oehrle.

The work reported here—across the various linguistic perspectives represented—still respects this view, which takes formalization as a useful accompaniment to linguistic theorizing. But in many respects, the place of logic, mathematics, and computation in this work goes deeper: subsequent developments have brought about a keener appreciation of how linguistic composition can be modeled as a form of grammatical deduction, how linguistic processing is essentially a form of computation, and how such ideas yield new and interesting connections between linguistic theory and practice, on the one hand, and results and theoretical developments in logic, in computation, and in the increasing connections between these two basic fields. Indeed, the trends in linguistic theory represented by the key words *constraints* and *resources* of our title correspond broadly to the most obvious duality in both logic and computation—namely, to the model-theoretic and the proof-theoretic perspectives, respectively, a distinction whose importance in the theory of computation has increased in time, as well. At the same time constraint-based and resource-sensitive theories of grammar share a lexicalist approach to linguistic theorizing. It is therefore fitting to include this volume in a book series that is intended to make available a broad range of linguistic, foundational and computational research in lexicalist linguistic frameworks.

Following the practice of G. J. Caesar and C. S. Peirce, we have divided the matter at hand into three parts. The first contains papers from the HPSG paradigm, papers whose formal focus falls on the formulation of linguistic constraints and the investigation of how suitably formulated constraints interact in the characterization of linguistic form and interpretation. The second part contains papers that draw on the proof-theoretical perspective in which the composition of linguistic structures is modeled as a form of resource-sensitive deduction in a grammatical logic. The third part contains formal investigations—investigations of grammatical structure and computation over grammatical structures. In the paragraphs that follow, we shall briefly describe the contents of each paper and situate it in the broader context of current linguistic theory.

Part I: HPSG

Bender and Flickinger's paper *Diachronic Evidence for Extended Argument Structure* presents a diachronic study of clauses in English that are headed by *as if*, *as though*, and *like*. They show that such clauses underwent a historical change from adjuncts to complements. This finding lends further support to recent proposals in HPSG to include certain adjuncts in the (extended) argument structure list of words. Bender and Flickinger argue that the inclusion of postverbal adjuncts in the argument structure of verbs provided the basis for the reanalysis of *as if*, *as though*, and *like* adjuncts as verbal complements. More generally, the paper exemplifies how the lexicalist perspective of HPSG, with highly-structured lexical representations motivated by synchronic considerations, can also provide an explanatory basis for certain types of syntactic change.

Bonami, Godard, and Marandin's contribution *Constituency and Word Order in French Subject Inversion* proposes an analysis in terms of *domain union* for French subject inversion in extraction contexts. In particular, they aim to provide an account for the fact that the subject may not only follow the VP, but also appear in between a verb and its complements. Their proposal steers a mid-course between a movement analysis based on extraposition of complements, and an analysis based on a flat sentential phrase structure in combination with *argument inheritance*. A domain-union analysis provides a phrase structure for subject-inverted sentences which is isomorphic to that of non-inverted sentences, thus avoiding a number of problems which arise in a flat phrase structure account. At the same time, domain union is more restricted than extraposition of complements, thus accounting for the bounded nature of inversion.

Crysmann's paper *Morphological Paradoxa in Fox* proposes a linearization-based HPSG account of complex verbs in Fox (Mesquakie), an Algonquian language spoken in Iowa. Complex verbs in Fox provide an interesting case study for the morphology-syntax interface. With respect to inflection and derivation they behave as single morphological units. At the same time, pre-verb and verb appear to be syntactically independent, since syntactically derived words can intervene between them. In order to account for their syntactic behavior Crysmann hypothesizes that complex verbs introduce more than one domain object (in the sense of Reape 1994 and Kathol 1995) into the syntax. Inflectional and derivational effects are then treated as morphological constraints on lists of domain objects.

Hinrichs and Nakazawa's contribution *VP Relatives in German* offers syntactic evidence in favor of a double-extraction analysis of German VP relatives: extraction of the entire relativized VP and extraction of the relative pronoun within the extracted VP. This double-extraction analysis is realized by two filler-head rules: the Filler-Head Relative Clause ID Rule, which licenses extraction of relativized VPs, and the Filler-Head Relative VP ID Rule, which is responsible for the internal extraction of relative pronouns within the VP. Following Sag's recent HPSG account of relative clauses in English (Sag, 1997), the authors treat the finite verb as the lexical head of the relative clause, thereby obviating the need for an empty relativizer (Pollard and Sag, 1994), which was mainly motivated as a placeholder for the semantic information contributed by the relative clause.

Kathol's paper *Restricted Modification and Polycategorial NPs in English* provides a critical evaluation of the conventional syntactic wisdom that restrictive nominal modifiers attach at the $\overline{\mathrm{N}}$-level. The conclusion is that, at least within a theory such as HPSG, there is little evidence for assuming $\overline{\mathrm{N}}$-attachment, and an analysis based on attachment at the NP-level is in fact to be preferred. He then turns to the broader issue of NP-internal structure in general, and develops a proposal in which both determiners and nouns are subtypes of *nominal*. This accounts for the fact that in some cases it seems most convenient to treat the noun as the head of NP, whereas in other cases the evidence points towards an analysis in which the determiner is the head.

Monachesi's contribution *The Syntactic Structure of Romanian Auxiliary (and Modal) Verbs* investigates the properties of Romanian auxiliaries. Auxiliaries in Romanian are always strictly adjacent to the main verb. In contrast to French or Italian, no adverb or subject can intervene between the two. This provides evidence for an analysis in which the auxiliary and main verb form a verbal complex. The distribution of pronominal clitics supports this analysis. Clitics which can be prefixed to their host, are always prefixed to the auxiliary—never to the main verb—a situation which can be accounted for by assuming that auxiliaries and verbs form a complex in which the auxiliary inherits the complements of the verb through *argument composition*. Modal verbs differ from auxiliaries in that adverbs may intervene between a modal and a subordinate verb. On the other hand, clitic climbing is possible. This suggests that modals do attract the arguments of their subordinate verbs through argument composition but that they do not head a verbal complex.

In Penn's paper *A Generalized-Domain-Based Approach to Serbo-Croatian Second Position Clitic Placement*, domain union plays a prominent role as well. Penn investigates second position clitics in Serbo-Croation, and argues that both prosodic and syntactic constituency constrain the possible placements of these clitics. He develops a theory of domain union which generalizes earlier work by Reape and Kathol. In particular, a notion of *compaction* of word order domains is proposed which is inspired by the theory of topological fields, and which can handle situations where prosodic and syntactic constituency are not isomorphic.

Van Eynde's paper *Major and Minor Pronouns in Dutch* proposes a distinction between major and minor categories. Major categories can act as the head of a non-vacuous phrasal projection, whereas minor categories cannot. In HPSG, the distinction can be implemented by assuming that the sort *cat* has *major* and *minor* as subsorts, and that only the former licences the valence features which are crucial for selecting syntactic dependents. Van Eynde applies his idea to Dutch full and reduced pronouns, and argues that the latter are minor categories. This accounts for the fact that reduced pronouns, in contrast to full pronouns, cannot be modified by adjectives, relatives, or appositions. Furthermore, a number of phenomena, such as the fact that reduced pronouns cannot be topicalized or conjoined may be accounted for in terms of this distinction. The paper concludes with the observation that the distinction also makes sense from a semantic point of view: only reduced pronouns in Dutch can act as expletives and can be interpreted generically.

Part II: Resource-sensitive approaches

Hendriks's paper *Contour and Structure: A Categorial Account of Alignment* achieves two notable goals. First, it integrates intonational aspects of topic and focus (in English) into a deductive approach to grammatical composition. Second, it shows how this perspective accommodates both lexical and phrasal constructions, thus resolving a theoretical conundrum of prosodic morphology. The focus of the paper is on two critical assumptions. The first is that both phonological words and syntactically composed phrases constitute *headed structures*, which can be modeled (following Moortgat and Morrill 1991) by a multi-modal system containing two independent versions of the non-associative Lambek calculus **NL**: one left-headed and one right-headed. The second is that intonational phenomena can be regarded as the consequence of a correspondence between proofs and phonological terms, exactly analogous to the familiar Curry-Howard correspondence between proofs and λ-terms. Hendriks concludes by showing how this basic system can be extended by modal operators and modally-sensitive type declarations to a system that accommodates a broader range of English intonational contours.

In *Topic and Focus Structures: The Dynamics of Tree Growth*, Kempson & Meyer-Viol advance a cross-linguistic typology of a variety of syntactic structures involving forms of topicalization or left-dislocation, based on simple and general assumptions about incremental processing and pronominal resolution. Incremental processing is modeled dynamically by the construction of tree-structures constrained by the order of lexical elements and the requirements on structure that they impose, as specified in a modal tree description language. Topic and focus structures have a natural characterization from this perspective, based on underspecified tree-descriptions. The different ways in which these underspecified representations can be updated in the course of incremental processing leads directly to the proposed typology of topic and focus structures, with additional consequences for the analysis of cross-over phenomena.

Moortgat's paper, *Constants of Grammatical Reasoning*, views grammatical composition from a deductive perspective and investigates the properties of an appropriate form of grammatical inference. Two basic questions that arises are: what are the constants, if any, of grammatical logic? and how are these constants compatible with the apparent structural variation that different natural languages display? To answer these questions, Moortgat draws on two converging logical traditions, Linear Logic and Lambek Calculus, which have focused attention on the impact of the 'structural rules' governing occurrences and multiplicities (Weakening and Contraction) and structural configurations (Permutation and Associativity), and shown how these operations can be locally controlled through modal operators. This leads Moortgat to a simple universal architecture for grammatical logic: the core is a family of unary and binary products; each product is associated with a corresponding family of implications, connected to the products by the residuation/adjointness laws which yield directly the familiar logical operations of *modus ponens*, the Deduction Theorem, and more. The lexicon of a grammar consists of a set of type declarations. Proofs can be interpreted according to the Curry-Howard correspondence, which associates each proof with a term of the typed λ-calculus in a way that has interesting consequences for the division of semantic labor between lexical specification and compositional syntactic structure. Finally, structural variation can be introduced by the addition of language-particular structural rules and interaction postulates (drawn from a universally available set). On this basis, Moortgat shows how the striking syntactic distinctions between Dutch and English extraction patterns can be elegantly characterized.

Oehrle's paper, *Binding as Term Rewriting*, investigates one way in which the general principles of dynamic semantics can be integrated with a multi-modal system of categorial grammar to provide an account of binding phenomena. Dynamic interpretation is based on two related properties: interpretation is context-dependent and context-affecting. Expressions can be characterized by their properties with respect to these two parameters; taking contexts to be structured objects makes this approach applicable to the variety of anaphoric expressions which have been the focus of binding theories in many different frameworks. Integration of this approach to binding with a system of grammatical inference requires that an account be given of how contexts are assembled and how communication between an expression and the context it depends upon and acts on is achieved. There are a number of ways to do this. The one developed in the paper uses a system of term rewriting, controlled via modal distinctions in the categorial system. The paper concludes with a discussion of how this approach extends to quantification, and an application of this extension to a puzzle involving quantification and cross-over.

Part III: Formal and Computational Issues

Kallmeyer's contribution *Synchronous Local TDGs and Scope Ambiguities* presents a formal account of the syntax-semantics interface for natural language grammars in the framework of *local Tree Description Grammars*, an extension of the Tree Adjoining Grammars. Syntactic and semantic representations are each formalized as underspecified tree descriptions in a local TDG. The two types of representations are linked by a synchronization relation between the two local TDGs. Tree descriptions in local TDGs allow an underspecification of the dominance relation between nodes in trees and thereby provide suitable underspecified representations for scope ambiguities. Since local TDGs are subject to a locality condition on derivations island constraints on quantifier scope follow naturally from the formalism and do not need to be stipulated. Kallmeyer illustrates the resulting treatment of scope ambiguities by a natural language fragment that includes quantifiers and relative clauses.

In *LIGs with Reduced Derivation Sets*, Michaelis and Wartena investigate a weakly context-free subclass of linear indexed grammar (LIG), called *unidirectional indexed grammar* (UIG), and its extended version. The linguistic interest in LIGs (and weaker forms thereof) is that they provide a simple model for investigating movement phenomena. The authors examplify this with WH-movement and Dutch cross-serial dependencies. More formally, in a LIG derivation each nonterminal has a stack of indices associated to it. This stack is passed on to exactly one daughter when the nonterminal gets rewritten. The authors investigate the consequences of restricting the set of possible candidates for inheritance of the index stack. For example, if the grammar only contains production rules that require the stack to be passed on following the rightmost branches only, then a right linear indexed grammar (RLIG) is obtained. RLIGs are weakly equivalent to context-free grammars, and provide an intuitively simple model for extraction in SVO-languages. Similarly, left linear indexed grammars (LLIGs) can be obtained. RLIG and LLIG can be combined into a UIG, in which the stack can be passed on either over the leftmost child or the rightmost child, with the restriction that direction can only be changed if the stack is empty. As a formal model, UIGs are closely related to the tree insertion grammars (TIGs) which are a kind of restricted tree adjoining grammars (TAGs). UIGs are still context-free, as the authors prove. The authors conclude the paper by constructing an extended version of UIGs, so-called EUIGs, which can be proven to have a weak generative capacity that is properly in-between CFLs and LILs. These grammars provide a formal model for SOV-languages. That way, the authors shed some light on the nature of non-local dependencies in natural languages.

Richter, Sailer, and Penn's paper *A Formal Interpretation of Relations and Quantification in HPSG* presents a formal language RSRL, a relational extension of the *Speciate Reentrant Logic* (SRL) introduced by King (1989). RSRL provides a formal semantics for the language introduced by Pollard and Sag to formalize HPSG grammars. The main innovations of RSRL, compared to SRL, is the treatment of relations and quantification that are used frequently in the statement of HPSG principles. Particularly noteworthy is the semantics of quantification that the authors develop. Quantification is restricted to components of a given linguistic object, rather than ranging over all or some objects in a given model.

Stabler's paper *Remnant Movement and Complexity* uses the phenomenon of 'remnant movement'—displacement of a constituent containing one or more traces of other displacements—to motivate a simple, spare formalism for grammars in the Minimalist framework. In this formalism, a grammar is just a pair $\langle Lex, \mathcal{F} \rangle$, with *Lex* a set of expressions and $\mathcal{F}$ a set of (partial) structure-building functions. In the grammars of interest here, lexical expressions consist of sequences of syntactic features associated with the phonological and semantic properties. And there are just two structure-building functions: *merge* and *move*, each with the interesting property that a pair of features of opposite polarity is canceled in each constructive step. Stabler shows how this formalism accommodates the seemingly complex Hungarian analyzed as an instance of remnant movement by Koopman and Szabolcsi (1998), and concludes with some observations involving the expressive power of the formalism, its affinities with other grammatical formalisms, and a revealing information-theoretic complexity measure that it supports.

Sources

The chapters of this book stem from papers originally presented at three recent and closely related conferences. In July 1997, the annual HPSG conference was held in Ithaca in conjunction with the LSA Summer Linguistics Institute at Cornell University. A month later, the third in a series of annual conferences on Formal Grammar took place in Aix-en-

Provence, during the weekend before the start of the ninth European Summer School in Logic, Language and Information (ESSLLI IX). When the organizers of these two conferences realized that they were planning to hold separate follow-up conferences in Saarbrücken in August 1998 in conjunction with ESSLLI X, they combined their plans and collaborated on the organization of the Joint Conference on Formal, Head-Driven, and Categorial Grammar (FHCG-98). Most of the papers included stem from the latter conference. The contributions by Hendriks, Kallmeyer, and Oehrle are based on presentations at the Formal Grammar conference, and the contributions by Hinrichs and Nakazawa and Penn are based on presentations at HPSG-97. A number of other papers derived from presentations at the latter conference are to be found in Volume 1 of the present series, *Lexical and Constructional Aspects of Linguistic Explanation* (Webelhuth et al., 1999).

Our goal in shaping this volume was to select high-quality papers whose content and exposition were such that they would be accessible to scholars both within and across the research traditions represented here and to a broader audience as well. The reviewing and editing process strongly favored papers that could be published without substantial revisions, thus allowing the publication of these papers relatively soon after the original conferences had taken place.

Acknowledgments

The editors would like to thank the authors for timely submission of their contributions, often on the basis of minimal instructions and LaTeX support from our side. We also thank the external reviewers for their cooperation and detailed comments on the submitted papers: their contribution to the result has been critical. Glyn Morrill has provided an essential impetus for the Formal Grammar meetings and we have benefited from his counsel in numerous ways. For assistance in proofreading the manuscript, we are grateful to Dr. Stephanie Schwarz.

Finally, we are indebted to the series editors, in particular Andreas Kathol, for assistance in the editing and reviewing process, and for arranging the publication of this volume within the *Studies in Constraint-Based Lexicalism* series. CSLI has provided us with LaTeX support in a variety of ways, and cooperated graciously and efficiently with our projected publication deadlines.

Gosse Bouma
Erhard Hinrichs
Geert-Jan Kruijff
Richard Oehrle

References

Chomsky, N. 1957. *Syntactic Structures*. The Hague: Mouton & Co.

Kathol, A. 1995. *Linearization-Based German Syntax*. Ph.D. thesis, Ohio State University.

King, P. 1989. *A Logical Formalism for Head-Driven Phrase Structure Grammar*. Ph.D. thesis, University of Manchester.

Koopman, H. and A. Szabolcsi. 1998. *Verbal Complexes*. UCLA. Forthcoming.

Moortgat, M. and G. Morrill. 1991. *Heads and Phrases: Type Calculus for Dependency and Constituent Structure*. OTS Research Paper. University of Utrecht: Research Institute for Language and Speech.

Pollard, C. J. and I. A. Sag. 1994. *Head-Driven Phrase Structure Grammar*. Chicago: University of Chicago Press.

Reape, M. 1994. Domain union and word order variation in German. In J. Nerbonne, K. Netter, and C. J. Pollard, eds., *German in Head-Driven Phrase Structure Grammar*, no. 46 in CSLI Lecture Notes, pages 151–197. Stanford University: CSLI Publications.

Sag, I. A. 1997. English relative clause constructions. *Journal of Linguistics* 33(2):431–483.

Webelhuth, G., J.-P. Koenig, and A. Kathol. 1999. *Lexical and Constructional Aspects of Linguistic Explanation*. Stanford: CSLI Publications. Distributed by Cambridge University Press.