Teilprojekt: A8

Thema: Operationale und deklarative Semantik grammatischer Systeme: Das Spannungsverhältnis zwischen Ableitung und logischer Charakterisierung linguistischer Strukturen

Leiter: Prof. Dr. Uwe Mönnich

Mitarbeiter: Tom Cornell, Ph.D.
Hans-Peter Kolb
Kai-Uwe Kühnberger
Jens Michaelis
Frank Morawietz
Craig Thiersch, Ph.D.

Inhalt

1 Kenntnisstand bei der letzten Antragstellung und Ausgangsfragestellung

After a long period of almost exclusively empirical work which characterized the Principles & Parameters (P&P, a.k.a. Government/Binding, GB) era of Generative Grammar research (Chomsky [9, 10]), with the advent of Minimalism (Chomsky [11]) formal language and complexity theoretic considerations have reentered—even though in a somewhat disguised form—the stage.

This seems to be a mixed blessing, however. While the P&P program is certainly not unproblematic from a formal point of view, its attempt to discover *substantive* properties of Natural Language has borne considerable fruit. Today's minimalist focus on the means of driving derivations is on first glance more amenable to—and more aware of the need for—formal analysis, yet it seems to be somewhat less successful in stimulating the empirical component of P&P research.

The two approaches illustrate very different attitudes towards formalization. To clarify the issue let us briefly consider the original research program of Generative Grammar as set forth in [8, 7]. Important details ignored and put in current terms, it rests on the conviction that

1. (a) there is a formal framework which allows exactly the specification of the class of Natural Languages: any theory (i.e., any consistent set of axioms) within this framework specifies a *potential* Natural Language;[1]
   (b) there exists a non-trivial set of axioms (the *initial state*, *Universal Grammar*), any *learnable extension* of which specifies a *possible* Natural Language, and every Natural Language has a theory which is a learnable extension of the initial state.

The goal of linguistic theorizing, then, is fourfold:

2. (a) to determine the primitive building blocks of linguistic theory (the *substantive universals*), e.g., the inventories of linguistic categories or features;
   (b) to determine the permissible operations and/or relations on these primitives;
   (c) to determine the initial state;
   (d) to specify a learning procedure (the *Language Acquisition Device, LAD*) which maps initial state and primary data onto the *steady state*, the theory (grammar) of a particular language.

[1] ... up to isomorphism, of course—after all, applications to totally unrelated domains cannot be ruled out.

Since the steady state is heavily under-determined by the primary data, (1b, 2c/d) comprises much of the gist of the Generative paradigm: a highly structured initial state reduces the burden on the LAD and increases the explanatory force of any grammar established this way—providing the underlying framework supports the establishment of consequences in a rigorous way. Hence, ideally, the formal framework is controlled to a considerable extent by substantive considerations such as questions of the structures admitted ("strong generative capacity"), or degrees of transparency in the expression of structural regularities.[2]

Conceptually, on the other hand, (1) and (2) are independent of each other. (2) can be understood as an entirely *empirical* agenda, to be explored under the very weak assumption that there exists a framework expressive enough to eventually formalize the results.

Yet, the advances—not the least through important contributions by Chomsky (e.g., [5, 6])—of Formal Language Theory in the 50s and 60s, which had provided much initial evidence for the feasibility of the basic approach, seemed to suggest a stronger rôle of the formal framework in Generative Grammar research. It had established, for the classes of the Chomsky hierarchy of formal languages, an intimate connection between the form of a grammar (the rule types employed), the computational resources needed to process it, and the type of (string) language specified by it. Applied to Natural Language this line of proceeding aims at establishing a formal system defined automata-theoretically and/or via rule types[3] which by its inherent restriction of expressive power provides guidance through the empirical mine-fields of linguistic theory formation.

Consequently, (1a) became a major concern of pre-P&P Generative theorizing, be it somewhat slanted towards considerations of computational complexity. Intricate—and not always accurate—arguments placed the natural languages on the Chomsky-Hierarchy somewhere between context-free and context-sensitive, and various varieties of transformational models were explored with the aim to discover the one "exactly right" formal framework, which, by our remarks above, would "automatically" provide for the right type of structures in a maximally transparent way.

A first seed of doubt about the adequacy of this approach was planted by Ross [61], who, based on careful analysis of a vast amount of empirical data, had demonstrated that many important regularities of Natural Language can only be expressed by means orthogonal to the descriptive apparatus of phrase-structural and transformational rules. Even though he had formulated his *constraints on variables in syntax* within the transformational framework, namely as restrictions on the interpretation of transformational rules, it was manifest that they were most naturally construed as substantive constraints on the structure of Natural Language.

---

[2]From the start demands for maximal transparency, elegance and conciseness in the expression of perceived structural regularities of Natural Language have played a major role in syntactic argumentation. Though not always well defined, they are considered criteria for an evaluation measure for grammars and can be considered the first economy principles.

[3]Questions of the relevance of commitments of this sort are at the heart of the recurring *representation* vs. *derivation* debate.

So when in the early 70s Peters and Ritchie formulated their famous (though not necessarily well-understood) result that there is an Aspects-style Transformational Grammar (TG) for any recursively enumerable language, i.e., that the formal framework of TG in itself does not only not restrict the class of natural languages in any way, but does not even ensure computability of the grammars couched in these terms [54, 55], this was the straw that broke the camel's back.[4] The original enthusiasm for questions of computational complexity was seriously abated, and the stage was set for a major shift in emphasis of linguistic theorizing towards *substantive* issues, i.e., from (1a) towards (2c): Not to get the formal framework "exactly right" became the primary goal now, but the discovery of "exactly the right" structural properties of Natural Language—in what ever way formulated.

It should be noted that the conceptional reorientation from *generating* to *licensing* theories is by no means restricted to the Chomskyan variety of Generative Grammar. The evolution of HPSG from its "generating" GPSG roots via the somewhat ambiguous approach of [56] to the clearly "licensing" based [57] is a case in point, as are recent developments in TAG theory towards the use of (partial) tree descriptions (cf., e.g., [59]). None of these revisions have lead to a comparable disintegration of the respective formal systems.[5]

However, except for lip-service, most of the little formal work that has been done on the basis of P&P theory has not been concerned with questions of (1) at all. Either formalization was restricted to some *fixed* version of the theory—usually in some "machine-readable" language under the heading of GB-parsing—without consideration for alternatives or future developments, or a strong meta-theory like full predicate logic—maybe even of higher order—was used in order to accommodate the wide variety of means of P&P theory formation.

Despite the indisputable merits of this work in elucidating many aspects of P&P theory neither strategy is likely to yield a formal framework for Generative Grammar in the sense of (1): The former, while relatively close to the ultimate purpose of the P&P model, is obviously premature: change seems to be about the only constant in linguistic theorizing and consequently there has not been a consistent, let alone complete, set of principles at any stage of GB-development. The latter, on the other hand, seems to go nicely with actual linguistic practice in that it provides all the flexibility needed to express new generalizations. However, in a Swiss-Army-Knife-type framework the particular structural properties of Natural Language grammars will always just be accidents without any explanatory force. Neither can such formalizations, due to their unrestrictive character, provide any significant guidance for linguistic theorizing, nor are there general decision procedures for such strong logics, hence these attempts are always in danger of formalizing for formalization's sake.

[4] Actually, their result had a second—and more important—part: Imposing one very natural constraint, i.e. the survivor property, a.k.a. *recoverability of deletion*, onto TG suffices to overcome the problem. However, to our knowledge the obvious route of trying to modify the formal system of TG so as to allow only grammars which have this property has never been taken.

[5] …nor, one might argue, to a comparable surge of empirical work. Whether there is some significant interdependence between these two facts would seem to warrant an investigation of its own.

The project advocates a different perspective, which warrants a more ambitious goal of the formal investigation of Natural Language syntax: The development of a *specific* meta-theory for the expression of theories of natural language syntax on the basis of a weak logic, which makes direct use of the primitive notions of grammatical theories. This line puts complexity theory and formal language theory firmly back on the map, although in a guise slightly different from the early days. In particular, it reconciles the rift between formal and substantive universals, providing a firm ground (i.e., clear formal bounds) for the formulation of substantive constraints.

2 Angewandte Methoden

Current linguistic theory has a number of prominent research traditions that use different grammar notations, and it is often unclear how the claims made in one tradition compare to claims in another. The difficulty in communicating results from one tradition to another, from one grammar notation to another, hampers the development of the field and leads to polemics based on misunderstandings. This project aims to find a neutral language in which principal results of any tradition can be stated, along with new results which this general perspective will reveal. For this purpose we draw from the mathematical tradition that has developed from looking at common properties of different formal systems in math: this is algebra. Algebra allows us to move from the mundane $2 + 3 = 3 + 2$ to the much more interesting $x + y = y + x$ (all x, y) – a very important and fundamental step in understanding mathematics. In a similar way, preliminary work by the two research groups involved in this study has shown that general and absolutely fundamental properties of human languages can be specified by moving to a just slightly more abstract framework, abstracting away from details of notation to get a clearer understanding of the fundamentals.

Languages are sometimes specified by "rewrite rules" which rewrite a sequence of categories and words until a sentence results, a sequence of lexical items to which no more rules apply. This way of specifying a language masks the fact that there are significant relations among the expressions of a language, relations that are easily made explicit just by shifting our perspective so that we regard the language as including a lexicon of basic elements ("words") which are then assembled by some basic structure building operations. In the latter algebraic approach, we have a (typically infinite) set of expressions with some relations defined on them by the rules for building complex expressions. Then the sentences are objects which are explicitly related to the parts that they contain, and to the larger expressions of which they are a part. Surprisingly, this latter perspective has not been given very much attention in the field until recently. In the first place, as shown by earlier work by Keenan and his collaborators, there is a natural algebraic notion of equivalence among expressions, equivalence with respect to "automorphisms" of the language, which allows us to notice several interesting things immediately, such as the existence of singleton equivalence classes, "fixed points" of the automorphisms, which correspond remarkably closely to what linguists have called "grammatical constants" or "function words." The further development of this perspective, proposed here, promises to be revealing.

The key concept for the formal treatment of linguistic theories of the licensing kind, on the other hand, is that of *descriptive complexity.*

While the *computational* complexity of a problem is usually defined in terms of the resources required for its computational solution on some machine model, *descriptive* complexity looks at the complexity of describing the problem (seen as a collection of relational structures) in a logic, measuring logical resources such as the type and number of variables, quantifiers, operators, etc. It is a non-trivial fact that there is a close correspondence between these two, with many natural logics corresponding exactly to independently defined complexity classes (cf. [23, 29]).

By hindsight, one can look upon Trakhtenbrot's theorem, according to which the set of all finitely satisfiable sentences of first-order logic is not decidable, as suggesting the problem of characterizing the cardinalities of finite models of first-order sentences. This problem, which has been coined the *spectrum* problem, was generalized by Fagin to the case of existential second-order, or Σ_1^1, sentences. In this context the problem consists in delineating *generalized spectra*, i.e., the classes of finite models of Σ_1^1 sentences.

It is easy to see that a spectrum is a decidable set of natural numbers. Given a finite signature τ and a natural number n, there are, up to isomorphism, only a finite number of τ-structures of cardinality n. For each of these structures, it takes only a finite amount of work to verify whether a first-order sentence ϕ of signature τ holds in it. By a similar procedure it can be shown that generalized spectra are decidable.

In order to relate logical expressibility with computational complexity finite structures have to be encoded as strings to turn them into possible inputs for Turing machines. Once a suitable convention for representing structures as strings has been chosen one can study the computational properties of finite structures. The remarkable result, known as Fagin's theorem, establishes the historically first equivalence between a problem in complexity theory and a problem in logical expressibility: Given a finite signature τ and a set of isomorphism-closed finite τ-structures K, K is a generalized spectrum if and only if K's set of encodings is in NP, where NP designates the class of problems that take nondeterministic polynomial time for a Turing machine to decide its individual instances.

This close connection between a syntactically restricted class of logical formulas and a time-restricted class of computations has attracted considerable attention both in the logic and the computer science community and was the beginning of a series of important results that provide analogous logical descriptions of other complexity classes. This research program, i.e. the systematic study and development of logical resources that capture complexity classes is well documented in the monograph by Ebbinghaus & Flum [18] to which we refer for further information. The basic aim of this program—*descriptive complexity theory*—consists in giving a characterization of the degree of complexity of a family of problems by looking into the structure of formal specifications in which these problems can be stated. In the words of Fagin [24], theorems that relate two distinct fields, logic and complexity theory in the present connection, are to be regarded as "mixed" theorems. They provide a precise justification for the use of methods from one field to solve problems in the other field, in the case at hand the employment of model-theoretic methods e.g., to attack problems in complexity theory.

As far as we can see, descriptive complexity theory in general has not exercised any direct influence on linguistic theorizing. The topic to which the contributions of this project belong and where from the computational perspective finite automata are taken into account rather than resource-bounded Turing machines has been called *descriptive theory of recognizability* in the recent overview by Thomas [69]. In terms of the logical specifications at issue, recognizability amounts to a restriction of existential second-order logic to its subsystem where only quantification of monadic second-order variables is permitted. We shall sketch in the following some of the remarkable achievements in this subfield of descriptive complexity theory that are directly related to central topics in theoretical linguistics. Along the way, we will try to point out in which way the interplay between computational and descriptive complexity reflects the change of paradigm from generating to licensing models in linguistics recounted in the first half of this introduction.

Before we start with this line of reasoning we have to insert a remark on the difference between two types of relations between syntax and semantics of formal systems. They can be illustrated with respect to Fagin's theorem. In one direction the theorem informs us that the question whether an arbitrary finite structure of signature τ belongs to the class of models of an existential second-order sentence of that signature can be decided in nondeterministic polynomial time. In other words, the theorem contains an upper bound for the complexity of the satisfaction relation of sentences in Σ_1^1. This satisfaction relation must not be confused with the satisfiability problem of theories stated in terms of the syntactic resources of a certain logic. The intuitive argument outlined above to the effect that the relation $\mathcal{A} \models \phi$ between a finite structure $\mathcal{A}$ and a first-order sentence ϕ is decidable is easily extended to a wide variety of logics. Fagin's theorem and subsequent results on the computational requirements of this relation lead to logical characterizations of a whole family of sub-recursive complexity classes. These results do not imply that the *theory* of an arbitrary structure or of a family of structures, whereby we designate by the theory of a structure—as is customary—the set of sentences satisfied by the (whole family of) structure(s), is a decidable set of sentences. Once the restriction to finite structures or the restriction to a single finite structure is lifted the question whether a sentence ϕ holds in an infinite structure or in a class of finite structures may become undecidable. It was shown in [58] that the monadic second-order theory of two successor functions $S2S$ is decidable. As soon as the syntactic means available for the formulation of the properties of the intended model of $S2S$, the infinite binary tree, are enriched by a single binary relation the resulting theory specified in terms of this enriched signature becomes undecidable. Even the restriction to the Σ_1^1 fragment of this extended theory suffices to obtain this negative result. This constitutes no contradiction with Fagin's result, as should be obvious. What it shows is the importance of the interplay between the class of models under consideration and the syntactic resources at hand in order to state their structural properties. One of the main attractions of descriptive complexity theory consists in this promising avenue it has opened for the systematic application of model-theoretic methods to problems of computational complexity.

In our discussion above of Chomsky's attitude towards issues in computational theory, we have emphasized the internal relation between fundamental tenets of the constraint-based variant of the Generative model and a pronounced distance towards questions of computational complexity. Under a certain perspective the considerations devoted a moment ago to the decisive role of the intended (class of) models can be read as a vindication of Chomsky's view defended in GB. Under this perspective the fact that a specific logical theory is undecidable should be no cause of concern as long as its model can be discarded for independent reasons having their origin in the nature of the human language faculty. Should, e.g., cognitivist assumptions of this sort require that the hypothesised principles that constrain this faculty of the mind, be only tested against finite structures, decidability becomes a non-issue. We shall return to this topic in connection with Rogers' formalization of the theory of Government/Binding.

The logical characterization of the complexity class NP was predated by a couple of results, especially the characterization of regular string and tree languages by means of monadic second-order logic of one ($S1S$) and multiple (SnS) successors, respectively. It is these earlier characterizations that provide the formal foundations for a renewed interest in logical approaches to grammar specifications. The work of the present project, while belonging to the logical tradition in linguistic theorizing, is particularly sensitive to the potential offered by logical techniques in generalizing formal language results from strings to other structures like trees and special types of finite graphs.

To be a little more specific, the original results in the descriptive theory of recognizability establish a tight connection between logical formalisms and language classes by providing translation procedures that transform logical specifications into finite automata equivalent to the language classes and vice versa. Büchi [4] and Elgot [19] have shown that regular string languages represented through finite (string) automata can be expressed by sentences in the weak monadic second-order logic with one successor. For tree languages an analogous result is well known to the effect that a tree language is definable in weak monadic second-order logic with multiple successors if and only if it is recognizable by a finite tree automaton [17, 67].

The logical approach to the specification of language classes involves a lot of advantageous properties that have paved the way to its application to linguistic issues. First, the equivalence between automata theoretic operational and logic oriented declarative formalisms leads to a lot of closure properties of the defined language classes that follow immediately from the closure of the specification logics with respect to the traditional operations like negation, conjunction, alternation and (universal and existential) quantification. Second, the transition from strings to finite model-theoretic structures of arbitrary signatures requires no extra conceptual or technical ideas in the logical framework whereas in formal language theory the step from string to tree languages and the concomitant distinction between weak and strong generative capacity constitutes a significant extension of the research agenda. Third, since the logical approach does not depend on an operational process which, starting from some given objects, generates its space of interpretation, but refers directly to an assumed universe of structures, its statements can be phrased in terms of linguistically significant notions that enter into universal principles and language-particular constraints. Finally, those logical languages that capture complexity classes indicate lower bounds on the computing resources a system has to make available that wants to use those structures that fall within the classes correlated with the corresponding logical language.

This spectrum of desirable properties constitutes by itself no guarantee for the success of a framework that avails itself of model-theoretic techniques for the study of linguistic structures. Languages that are expressive enough to define significant properties of natural languages may be too powerful from the perspective of complexity theory or even classical recursion theory. On the other hand, logics which are well-behaved in terms of the language families definable by them may lack the syntactic resources necessary for the direct determination of relevant linguistic notions and principles.

Given this threat of a possible failure between an overpowering Scylla and an impoverished Charybdis, the value of the formalization of major parts of Head-Driven Phrase Structure Grammar achieved in Paul King's Speciate Re-entrant Logic [35, 37] and of the formalization of essential aspects of GB theory attained within a definitional extension of monadic second-order logic of multiple successors in Rogers [60] cannot be overestimated. The intended models of King's logic being a special type of graphs with labeled edges, we restrict our attention in the following on Rogers' work [60], because he is concerned with another type of graphs, category-labeled finite trees, that have been the focus of interest in the Generative tradition and that constitute the preferred "domain of discourse" of the papers assembled in the context of the present project.

As was pointed out above, the weak variant of the logic chosen by Rogers is severely restricted from the point of view of descriptive complexity theory. Languages are definable model-theoretically in this logic just in case they are an instance of a recognizable tree language, where we disregard in the context of the present discussion the issue of unbounded finite out-degree of nonterminal nodes in trees belonging to such a language. In spite of this low degree of computing resources to decide membership in a language specified by means of this monadic second-order logic the logic is surprisingly flexible in capturing central concepts of the P&P approach. This apparent conflict can be easily accounted for by looking at particular definitions proposed in Rogers' work. It then becomes immediately clear that two main ingredients underlie the success of this attempt to couch a linguistic theory of the licensing type in a precise logical medium. There is, on the one hand, the device of second-order quantification which allows one to speak about arbitrary sets of nodes in a tree model. While this device implies already a great advantage in comparison with the definitional power offered by first-order logic, it would by itself carry not very far without the possibility of addressing directly linguistically significant tree configurations. The (extended) signature of the logic contains relations like dominance, immediate dominance, and precedence, which provide the foundations for a quasi-effective translation procedure from the technical jargon familiar to linguists into the specific second-order logic at hand.

Both approaches share a common drawback because it turns out that once these logical approaches are slightly extended either on the formal or on the conceptual side they become undecidable. Rogers shows that enriching his logic by a relation adequate for a representation of free indexing leads to its undecidability. King et al. [38] show that a plausible notion of grammaticality involves undecidability for King's [35] system. When looking at the actual proofs of those results it turns out that they do not concern the satisfiability problem for a class K, where K consists of a single finite structure or of a finite set of such structures, but where K is either an infinite structure or an infinite set of finite structures fulfilling certain closure conditions. This situation suggests the interesting question whether the context in which a native speaker learns or uses his or her language is adequately described by confronting the language module with a surveyable finite input or whether the topic of foremost interest to the linguist is the decidability of the representational medium adduced to give a precise characterization of the structure of that central cognitive faculty the language module incorporates. We have to leave this question open and hope to return to it on another occasion.

3 Ergebnisse und ihre Bedeutung

There are many kinds of structural phenomena in natural languages that cannot be captured by context-free string grammars or regular tree grammars. On the other hand, all the constructions that have been cited as evidence for the need of assuming context-sensitive grammatical devices for the description of natural languages seem to be amenable to an analysis within the framework of context-free tree grammars. Based on the Mezei-Wright [43] result according to which structural accounts of the context-free tree level can be lifted to the regular tree level where composition and projection occur as explicit node labels, work of the present project has focused on a fine-grained analysis of the process connecting the initial level of explicit trees with the intended level of context-free or macro trees. In accordance with the three types of classical approaches in formal language theory we have provided logical, grammatical and automata-theoretic characterizations of the homomorphism relating the initial tree algebra of Lawvere terms to the substitution algebra of macro terms. Along the way first steps were taken in the direction of "reverse" linguistics. For contemporary models such as Tree Adjoining Grammar and Minimalism it has been shown how to accommodate their main ideas within the algebraic framework outlined above.

Getting beyond these preliminary observations, it is interesting to see what the prominent existing approaches in linguistic theory look like when recast in this general setting. Work on this topic by the project has shown that a wide range of proposals can be formulated in an algebra of trees, including central proposals in "tree adjoining grammar" (TAG), "categorial grammar" (CG), and even simple formulations of Chomskian "minimalist grammars" (MG) that have been formalized by the work of Stabler, Michaelis and others involved in this project. This work has led to precise ways of translating or "embedding" the proposals in one notation into proposals of other notations, allowing comparisons among theories that were not possible before. In addition, the same intertranslatability results has allowed parsing algorithms for one notation to be adapted to other notations. Already, the embedding result of Michaelis has led to the first general parsing procedure for minimalist grammars, in the work of Stabler and Harkema [64, 28]. Further development of this perspective promises much deeper insight into the fundamental properties of human languages, properties that notational diversity is too often obscuring. In particular, ongoing work is revealing that a very wide range of notations are intertranslatable, and with this development, the prospects for clear, precise and general statements of the universal properties of human languages look very good, including in particular, specifications of minimal sets of basic operations required to define human languages.

The starting point in the Tübingen context was [49]. That paper deals with the problem of how to account for mildly context-sensitive phenomena with the help of the monadic second-order logic with multiple successors (MSO). The limited expressive power of MSO in its original set-up makes it impossible to formulate the solution in a direct way but using techniques from universal algebra one can transform these phenomena into a slightly different form whereby they become context-free and as such definable in tree logic.

Mönnich uses an "algebraicized" variant of grammars with macro-like productions [25] to present an analysis of structural dependencies which are beyond the reach of context-free devices. Based on the technical device of a derived or lifted alphabet the macro-like productions are transformed into the familiar productions of a regular tree grammar. In informal terms, the formation of a derived alphabet can be explained as a nominalization process turning functions into objects. The language generated by this nominalized procedure is context-free and as such amenable to logical analysis in the framework of monadic second-order logic with multiple successors.

During the translation of the original alphabet into its nominalized derived counterpart new operators and constants are introduced that change the configurational shape of the translated expressions, which leads to a certain tension between the definitions for linguistic concepts in terms of structural configurations of the original trees and their counterparts in the lifted context. Towards the end of the paper a reduction is sketched, itself defined in MSO, which provides for a translation of the lifted tree structures into the intended ones.

Mönnich's result that non-recognizable sets of structures can—up to homomorphism—be coded context-freely, thereby returning them into the scope of a MSO treatment, is the starting point of [39], which tries to relate the structures obtained from a grammar over the nominalized derived alphabet directly to the familiar tree structures of linguistic theory. Using a simplified minimalist analysis of the Germanic verb complex as the primary illustrative example it explores the feasibility and identifies the respective problems of trying to establish in a rigorous and effective way a direct relation (a) between a linguistic "grammar" and the set of structures generated by a "lifted" equivalent tree grammar, and (b) between the class of linguistic theories (in the logical sense) and their MSO counterparts over the derived alphabet. (1) amounts to redefining the linguistic primitives (in MSO) systematically in such a way that it is possible to "forget" the additional structure when talking (in MSO) about the linguistic structural relations holding in a given tree, while (2) involves a direct interpretation *in the linguistic theory* of the operative function symbols of the macro tree grammars, and of the composition and projection symbols of their regular encodings. In particular an interpretation which relates the additional symbols to the derivational ingredients of a minimalist theory would allow us to state in MSO facts not just about the linguistic structures, but also about the way they are arrived at, thereby obtaining the means for a formal investigation of the resources of modern Generative Grammar itself.

In the case of (a) an effective syntactic interpretation of the intended trees in the explicit ones can be given at least for relations based on the central notion of *dominance*, but there is no direct solution for (b) since, contrary to appearance, the explicit structures do not encode movement in any way. However, yet another explicit encoding based on T-markers may provide a way out of this dilemma.

The work by Kolb and Mönnich which is centered around the operational notion of a tree walking automaton and the algebraic notion of a uniquely given homomorphism is considerable extended in the following papers: [41, 40, 46].

The main result of [41] is a description of cross-serial dependencies equivalently in terms of regular tree languages (to express the recursive properties) and regular string languages (to establish the intended linguistic relations) and transductions definable in monadic second-order (MSO) logic which are defined on a domain of finite trees that is characterized as the model set of a closed MSO formula. In fact, the paper does not deal directly with the two types of regular language families and MSO definable structures, but instead considers implementation in the form of finite-state (tree) automata to emphasize the effectivity of the approach.

[40] gives an additional description of cross-serial dependencies in terms of a special type of tree transformations effected by a linear non-deleting macro tree transducer. To be more precise, the paper constructs, in a first step, a regular representation of the cross-serial dependencies using a certain amount of explicit control information. Trees exhibiting this control information are elements of an absolutely free algebra on a signature that results from the well-known derivation process explained above. The macro tree transducer (MTT) serves to evaluate the tree terms of the free algebra in a different semantic domain. MTTs are a powerful syntax-directed translation device in which the translation of an input tree not only depends on its subtrees but also on the context. The particular MTT constructed in the paper takes the tree language generated by the regular tree grammar referred to above as its input and transduces it into the intended structures that are specified by the logical MSO transduction. In other words, the transduction constitutes the implementation of the uniquely given homomorphism which sends the trees with the explicit control information into the originally intended structures.

The instruction set of the MTT is universal and does not depend on the specification of the particular regular tree family to be transformed. Its rules serve to reverse the lifting process by evaluating the instructions implied in the composition and projection symbols.

[46] extends the previous work by the first author [44] which shows how to encode an arbitrary minimalist grammar (MG) in the sense of Stabler [63] (and thus in the sense of Stabler [64], the chain-based incarnation of the former) into a weakly equivalent linear context-free rewriting system (LCFRS). The paper presents a translation from the formalism of multiple context-free grammars (MCFGs)—a weakly equivalent extension of LCFRSs—into regular tree grammars (RTGs). The idea behind the translation is to "lift" the MCFG-rules to RTG-rules by viewing them as Lawvere terms. Furthermore, the paper uses the equivalence of RTGs, monadic second-order (MSO) logic (on trees) and tree automata to give an algebraic and a logical description of the lifted trees. Since the trees characterized by an RTG contain additional "non-linguistic" information they are then unpacked with a monadic second-order (MSO) transduction thereby

giving both an operational and a denotational description of the tree sets involved. The MSO transduction is built upon a tree-walking automaton (with tests).

In addition to the result just mentioned, the paper presents a further implementation of the MSO transduction by another simple MTT. It is worth mentioning that the tree transduction is given in terms of the "lifted" terms of the Lawvere algebra. It would have been possible to turn the MCFG directly into an equivalent MTT because in the paper it is also shown that MCFGs are strongly equivalent with a family of languages that are generated by a particularly simple form of attribute grammars (AGs). AGs are a much more powerful device than MCFGs and can easily accommodate phenomena of natural languages that are beyond the reach of the mildly context-sensitive grammar formalisms just enumerated. Analyzing multiple context-free grammars (MCFGs) as a special type of AG makes for a smooth transition into higher echelons of a refined Chomsky hierarchy without leaving the conceptual framework of AGs that has been successfully applied in such diverse areas as compiler construction and recursive program schemes.

Besides its analysis of multiple context-free grammars, the previous paper also constitutes the first example of what was called above "reverse" linguistics. Close inspection of the construction in [44] reveals that for the set of trees appearing as intermediate steps in converging derivations of a minimalist grammar, one can define a finite partition. The equivalence classes of this partition are formed by sets of trees where the features triggering movement appear in identical structural positions. Each nonterminal in a corresponding MCFG represents such an equivalence class, i.e., an infinite set of trees.

A further example is provided by [47] which presents tree adjoining grammars as monadic context-free tree grammars. The central notion in this algebraic perspective on tree adjoining grammars is that of a higher-order substitution. The greatest advantage that comes with such a presentation derives from its close relationship with the algebraic, logical and automata-theoretic techniques that were put to use in the papers summarized above. The algebraic perspective thus not only adds another characterization of tree adjoining languages to the already long list of equivalences with restrictive production systems, but it also makes available the whole gamut of transductions between initial and intended semantic structures that are at the core of the algebraic approach to linguistic theory.

Some of the preceeding results rely on the fact that Minimalist Grammars in the sense of Stabler [63, 64] fall indeed in the range of mildly context-sensitive grammar formalisms, and indeed they are weakly equivalent. This result is proven in [45] via a translation into LCFRSs. It follows immediately that the class of languages generated by Minimalist Grammars forms an abstract family of languages. Additionally, the proof solves the question posed by Stabler whether the two versions of Minimalist Grammars proposed in [63] and [64] are weakly equivalent.

The project has also investigated the applicability of the theoretical results by Doner, Thatcher & Wright and Rogers on a different level. The effectiveness of the translation from formulas to finite state devices—tree automata or transducer—enables experimenting with the proposed linguistic constraints in a practical, parsing oriented setting, see [2, 15, 50, 51, 52, 53].

In this sense Morawietz & Cornell [53] discuss possible applications of a system which uses automata-based theorem-proving techniques drawing on the decidability proof for weak MSO logic on trees to implement linguistic processing and theory verification. Despite a staggering complexity bound, the success of and the continuing work on these techniques in computer science promises a usable tool to test formalizations of grammars. The advantages are readily apparent. The direct use of a succinct and flexible description language together with an environment to test the formalizations with the resulting finite, deterministic tree automata offers a way of combining the needs of both formalization and processing. The contribution of this paper is threefold. Firstly it is shown how to use the technique for the verification of separate modules of a P&P grammar and secondly for the approximation of an entire P&P theory. And thirdly, the language of the MSO tree logic is extended with recursively definable predicates to overcome remaining engineering problems. Cornell [15] forms the basis of the preceeding work in the sense that is shows that tree automata can be used for parsing and that this process is really just a variant of the standardly know deductive parsing techniques.

Morawietz, in a series of papers [2, 50, 51, 52], has investigated the usability of techniques from deductive parsing in a constraint programming setting, thus enabling both paradigms to coexist in a theoretically clean setup by extending the proposal in Shieber et al. [62] which provides a meta interpreter for several deduction systems. The papers show how to directly realize the inference rules for chart parsers as Constraint Handling Rules (Frühwirt [26]) by viewing the items of a conventional chart parser as constraints and the constraint base as a chart. This allows the direct use of the constraint resolution process to parse sentences.

Cornell, again in a couple of papers [13, 14], has discussed the confluence of denotational and operational views of the Minimalist program in a type logical setting. Many researchers in the past few years have remarked on the apparent relation between ideas from Chomsky's minimalist program and fundamental ideas in categorial grammar. The main feature of minimalist derivations which has suggested a connection with categorial grammar is the treatment of feature checking in Chapter 4 of Chomsky (1995). The idea that a feature on a moving phrase might be canceled against a feature on the structure to which it moves is very reminiscent of the basic cancellation rules of categorial grammar. Seen in this light, the derivational need to cancel features reduces to the proof-theoretic need to reduce a hypothesis to axioms. This paper represents a first stab at the problem of developing ideas from the minimalist program within a type-logical deductive grammar framework. Some positive results indicate that the representational approach to minimalism of Brody [3], involving crucially the idea of "presyntactic chain formation", can be formally related to a more strictly derivational approach by using a labeling algebra which simply reflects the phrase structure tree underlying the feature-driven derivation. A surprising negative result suggests that, given Chomsky's assumptions about the structure of the initial numeration, the generalized transformation Merge may be harder to implement in this setting than Move.

4 Vergleiche mit Arbeiten ausserhalb des SFB

The group led by Ed Keenan in Los Angeles has undertaken several foundational studies, in which it was observed, first, that "equivalence with respect to automorphisms of the language" provides a precise version of the linguists' intuitions about the notion of "having the same structure" [32, 33, 34]. Considering a range of grammars for (fragments of) human languages, it is striking that the "grammatical morphemes" or "closed class functional categories" are typically not equivalent to anything else: that is, these elements have unique roles in the language. So it appears that the size of the sets of equivalence classes of expressions corresponds to distinctions that linguists have been considering for decades, though appropriate ways of specifying these elements has been lacking (e.g., [20]. This setting also provides an appropriate setting for characterizing appropriate relations between syntax and semantics [33, 30, 27], and we notice that, at least to a remarkably good first approximation, grammatical constants denote "semantic constants," where by this latter notion we have in mind the well-known Tarskian idea of being fixed by the automorphisms of the model [31]. In a preliminary survey of some other universal claims about human languages, Stabler & Keenan [66] identify some respects in which the standard equivalences with respect to isomorphisms, or identity of symmetry groups, are too "brittle," and their first proposals for slightly more "stable" notions are being explored in ongoing work.

Stabler himself has presented the first formalizations of Chomsky's Minimalist Program [12] in a growing number of papers [63, 65, 64, 66] which influenced and were influenced by the collaboration with members of the project. In these papers Stabler not only argues for the linguistic adequacy of his proposed formalism but also presents several questions concerning their formal complexity. The work by Michaelis and by Michaelis, Mönnich and Morawietz together with Stabler led to the proof sketched above that Minimalist Grammars are indeed mildly context-sensitive grammar formalisms and in turn changed the formal presentation of Minimalist Grammars.

Outside linguistics proper, recent developments in theoretical computer science that regard derivation trees of various tree and graph grammars as elements of a regular tree language are related to the general approach pursued in the present project. This perspective on derivation trees was first introduced in a paper by Mezei & Wright [43] where regular tree languages are used to define subsets of arbitrary algebras in the following way. For a ranked set of operators, a Σ-algebra is a carrier A on which an operation $\sigma_A : A^k \longrightarrow A$ is specified for every $\sigma \in \Sigma$ of rank k. Every tree $t \in T_\Sigma$ can then be seen as an expression that denotes an element $val(t)$ in A, where val designates the unique homomorphism from the free Σ-algebra T_Σ into A. In this sense, a regular tree grammar that generates a family of trees in T_Σ may be regarded as specifying a subset of S by taking the values of the generated trees in A. It is well known that context-free languages are the "regular" subsets of the algebra of strings with concatenation as the only operation. For graph languages, this approach was generalized by Bauderon & Courcelle [1].

If one considers the algebra A of all (hyper)graphs with any possible production as operation, then several types of context-free graph grammars can be expressed as systems of equations over appropriate sets of graph expressions. Since these operations constitute fairly intricate substitution operations one major research goal in this field has been the isolation of a primitive set of simple operations that are functionally complete with respect to the full set of graph productions. Bauderon & Courcelle were able to show that such a simple set exists for the major families of graph grammars. An analogous result, which, actually, was proven earlier by Engelfriet and Schmidt [21, 22], holds for context-free tree grammars. In this case one can restrict oneself to the two operations of tree composition and projection. Several papers of the present project use this simple set of tree operations as their starting point. In some cases it has turned out to be useful to embed the approach followed by the project with the even more general framework of algebraic theories as conceived by Lawvere [42]. The descriptions of the tree adjoining languages and of the Minimalist syntax would have become far more opaque without this elegant part of category theory.

5 Offene Fragen

Given the background of the general research agenda specified in the introduction, four different areas of further research impose themselves that seem to be well suited to continue the work described in the previous sections. These areas are arranged around issues of foundational, applied, linguistic and methodological character.

Foundation.

Even though the existence of context-sensitive phenomena in natural languages is firmly established there is an open debate as to the exact level of complexity that is to be assumed for a descriptively adequate linguistic theory. To be more specific, there is an empirical issue about the status of cases of so-called Suffixaufnahme. If these instances exhibit the kind of productivity that is claimed for the phenomena of case-stacking attested in Old Georgian mildly context-sensitive devices are not powerful enough to accommodate these structures. This is true in particular of the family of multiple context-free grammars whose Parikh image is semi-linear. The condition in the definition of multiple context-free grammars that is responsible for this weakness is the stipulation that parameters have at most one occurrence in the left-hand side of productions. Dropping this condition leads to a much larger class of languages where structures like unlimited case-stackings in Old Georgian can easily be described. A major research issue would therefore be tied to the question of whether natural languages are semi-linear. Even a negative answer would not constitute a refutation of universal algebraic approach of the introduction since it has been shown in [48] that the generative power provided by context-free tree grammars is enough for the Old Georgian case structures.

Methodology.

In the overview of results achieved so far by the applicants it has been pointed out that we hope to be able to give a faithful representation of leading contemporary linguistic models within our general algebraic framework. The case of Minimalism has been covered in the review of [46] and will not be repeated here. The first steps in the direction of tree adjoining grammars were also reported upon in the preceeding paragraphs. It is hypothesized that these steps lay the groundwork for an account of tree adjoining grammars along the lines of our algebraic approach. What remains to be done is the construction of simple macro tree transducers, tree walking automata and MSO transductions that would transform the control information in explicit trees into the structures intended by the proponents of the tree adjoining model. A successful completion of this task would also provide the necessary prerequisites for an equational analysis of tree adjoining languages according to the denotational paradigm. A similar representation of licensing theories seems to be much more difficult. As is well known, the celebrated result by [16, 68] on the equivalence between logical and grammatical specifications of regular families of trees cannot be extended to graphs. However, a set of graphs that is second-order definable is presentable as a regular set of trees over a special signature of graph operations. This latter regular set is again amenable to the kind of transformation that is characteristic for our approach. The feasibility of a successful re-analysis of a prominent member of the licensing paradigm like Head-Driven Phrase Structure Grammar is therefore predicated upon a specification of graph-like feature structures by means of monadic second-order logic. Our conjecture is that such a specification can be done by means of a meaning-preserving translation of King's speciate re-entrant logic [36] into monadic second-order logic. As King's proposal has become the accepted logical support for HPSG formalisms such a translation into a more standard system such as MSO logic would at the same time pave the way for an investigation of the formal properties of licensing models by means of established mathematical techniques. We will also model licensing theories and their underlying graph-like data structures with coalgebraic methods. These methods are particularly well suited for modeling graphs and should therefore allow for a direct way of modeling licensing theories.

Application.

The general algebraic approach as a unifying framework is not only of considerable theoretical appeal but can also be put to use in areas of a more practical interest. We restrict ourselves to two subjects of immediate relevance to the proposal: parsing and web technology. Given the two-step approach outlined in section 2 on methodology, a similarly conceived division of labor is applicable to create a powerful parsing mechanism. As finite-state tree automata are neutral with respect to the difference between generating and accepting devices there are no theoretical or practical obstacles concerning the level of regular tree sets. Since these trees do not constitute the intended linguistically significant structures they have to be intersected with a further finite-state device to filter out the intended results on the transformed level. Preliminary work by the project shows that this approach promises to run effectively and can be integrated with related advances in parsing methodology achieved by the Los Angeles group. Concerning the most recent developments in designing tools for creating dynamic web pages the mark-up language XML emerges as the lingua franca of electronic content. XML-based document type definitions (DTD) have also become the standard format for the information modeling of linguistic field data. As it turns out DTDs correspond to regular tree languages with non-local information. As has been stated in the previous section, parameters in macro tree transducers are intended to handle certain types of context information. Cursory inspection of the kind of non-local information needed for the transformation of DTDs suggests that it can again be accommodated within the general two-tiers algebraic framework that represents the unifying theme of the present proposal.

Linguistic Universals.

It would be nice if the set of grammatical functions characteristic for natural languages could be reduced to a small primitive set. To recall, we regard languages as the closure of their lexica with respect to their grammatical functions. These functions fall into two disjoint groups, book-keeping functions like composition, projection and tuple-formation, on the one hand, and linguistic functions proper like case-marking and other inflexional processes. The book-keeping functions seem to be mainly responsible for the recursive aspect of language whereas the linguistic functions seem to play a special role in assessing the structural similarity between languages. The algebraic approach towards language invariants developed by the Los Angeles group has already been applied to several fragments of languages and has helped to put the comparison between different linguistic similarity types on a firm methodological basis [33, 34, 66]. The remaining members of the project hope to be able to pursue this work and take further material from unrelated languages into consideration. The overall aim is the establishment of a list of primitive grammatical operations, both in the book-keeping and in the linguistic class, that can then be tested against additional data from diverse language families.

6 Veröffentlichungen und Vorträge

Veröffentlichungen

Blache, P. und Morawietz, F. *A Non-Generative Contraint-Based Formalism* (im Arbeitsbericht des SFB 340 150, 2000).

Cornell, T. *Island Effects in Type Logical Approaches to the Minimalist Program* (Proceedings of the Formal, Head-Driven and Categorial Grammar Conference, Saarbrücken, 1998).

Cornell, T. *Representational Minimalism* (The Mathematics of Syntactic Structure, Studies in Generative Grammar 44, Mouton de Gruyter, 1999).

Cornell, T. *Derivational and Representational Views of Minimalist Transformational Grammar* (in den Proceedings der LACL Konferenz, LNAI 1582, Springer, 1999).

Cornell, T. *Parsing and Grammar engineering with Tree Automata* (in den Proceedings des Workshops AMILP 2000 in Iowa, 2000).

Kolb, H.-P., Mönnich, U. und Morawietz, F. *Regular Description of Cross-Serial Dependencies* (in den Proceedings der MOL 6 Konferenz, 1999).

Kolb, H.-P. und Mönnich, U. *The Mathematics of Syntactic Structure: Trees and their Logics* (Studies in Generative Grammar 44, Mouton de Gruyter, 1999).

Kolb, H.-P. *Macros for Minimalism? Towards Weak Descriptions for Strong Structures* (The Mathematics of Syntactic Structure, Studies in Generative Grammar 44, Mouton de Gruyter, 1999).

Kolb, H.-P., Mönnich, U. und Morawietz, F. *Logical Description of Cross-Serial Dependencies* ([extended abstract] in den Proceedings des Workshops *Finite Model Theory and its Applications* der Konferenz *FLoC '99*, 1999).

Kolb, H.-P., Mönnich, U. und Morawietz, F. *Descriptions of Cross-Serial Dependencies* (in *Grammars*, 2000).

Kolb, H.-P., Michaelis, J., Mönnich, U. und Morawietz, F. *An Operational and Denotational Approach to Non-Context-Freeness* (erscheint in Theoretical Computer Science).

Kühnberger, K.-U. *Frameworks for Circular Phenomena. Possibilities of Modeling Pathological Expressions in Formal and Natural Languages*, Dissertation, 2001.

Kühnberger, K.-U. *Remarks on the Complexity of Truth* (in den Proceedings von Foundations of the Formal Sciences II, in press).

Michaelis, J., Mönnich, U. und Morawietz, F. *Algebraic Description of Derivational Minimalism* (in den Proceedings des Workshops AMILP 2000 in Iowa, 2000).

Michaelis, J., Mönnich, U. und Morawietz, F. *Derivational Minimalism in Two Regular and Logical Steps* (in den Proceedings des Workshops TAG+5 in Paris, 2000).

Michaelis, J., Mönnich, U. und Morawietz, F. *On Minimalist Attribute Grammars and Macro Tree Transducers* (erscheint in einem CSLI Band zur internationalen Tagung des SFB 340 *Linguistic Form and its Computation* in Bad Teinach).

Michaelis, J. *On Formal Properties of Minimalist Grammars*, Dissertation, 2001.

Mönnich, U. *TAGs M-constructed* (TAG+ 4th Workshop, Philadelphia, 1998).

Mönnich, U. *On Cloning Contextfreeness* (The Mathematics of Syntactic Structure, Studies in Generative Grammar 44, Mouton de Gruyter, 1999).

Morawietz, F. und Cornell, T. *The Logic-Automaton Connection in Linguistics* (in den Proceedings der LACL Konferenz, LNAI 1582, Springer, 1999).

Morawietz, F. *MSO Logic, Tree Automata and CLP* (in: H.-P. Kolb und U. Mönnich (Hrsg) *The Mathematics of Syntactic Structure. Trees and their Logics*, Mouton de Gruyter, Studies in Generative Grammar 44, 1999).

Morawietz, F. *Chart Parsing as Constraint Propagation* (in den Proceedings des IWPT 2000).

Morawietz, F. *Chart Parsing and Constraint Programming* (in den Proceedings der Konferenz Coling 2000).

Morawietz, F. *Some Aspects of Natural Language Processing and Constraint Programming* (Editor, Arbeitsberichte des SFB 340 150, 2000).

Morawietz, F. *Chart Parsing as Contraint Propagation* (im Arbeitsbericht des SFB 340 150, 2000).

Vorträge und sonstige Aktivitäten

Morawietz, F. *The Logic-Automaton Connection*, (Workshop *Applications of Constraint-Based Programming to Computational Linguistics*, SFB 340, Blaubeuren 2–6. Mai 1998)

Cornell, T. *Island Effects in Type Logical Approaches to the Minimalist Program* (*Formal, Head-Driven and Categorial Grammar Conference*, Saarbrücken, August 1998)

Mönnich, U. *TAGs M-constructed* (Tag+4, UPenn, Philadelphia, August 1998)

Cornell, T. *Feature-Checking Nets: Graphical Representations of Minimalist Derivations*, (Innovationskolleg *Formale Modelle kognitiver Komplexität* Universität Potsdam, Workshop *Grammars for Linguistic Structures*, Universität Potsdam, Campus Golm, 23./24. Okt. 98)

Kolb, H.-P. *Macros for Minimalism?*, (Innovationskolleg *Formale Modelle kognitiver Komplexität* Universität Potsdam, Workshop *Grammars for Linguistic Structures*, Universität Potsdam, Campus Golm, 23./24. Okt. 98)

Mönnich, U. *TAGs M-Constructed*, (Innovationskolleg *Formale Modelle kognitiver Komplexität* Universität Potsdam, Workshop *Grammars for Linguistic Structures*, Universität Potsdam, Campus Golm, 23./24. Okt. 98)

Morawietz, F. *Compiling MSO Logic into Tree Automata*, (Innovationskolleg *Formale Modelle kognitiver Komplexität* Universität Potsdam, Workshop *Grammars for Linguistic Structures*, Universität Potsdam, Campus Golm, 23./24. Okt. 98)

Morawietz, F. und Kolb, H.-P. *Der Weg ist das Ziel: Automaten und Kreuzabhängigkeiten*, (SFB 340 Konferenz, Stuttgart, Jan. 1999)

Morawietz, F. *A Home for the Homepageless*, (Kolloquium des Graduiertenkollegs *Integriertes Linguistikstudium*, Universität Tübingen, Feb. 1999)

Morawietz, F. *MSO Logic for Natural Language Processing*, (Laboratoire Parole et Langage, CNRS & Université de Provence, Apr. 1999)

Cornell, T. *Derivation and Representation from a Categorial Perspective* (Workshop *Komplexitätstheoretische Eigenschaften von Grammatikformalismen*, Tübingen, Jun. 1999)

Mönnich, U. und Morawietz, F. *Descriptions of Cross-Serial Dependencies* (Workshop *Komplexitätstheoretische Eigenschaften von Grammatikformalismen*, Tübingen, Jun. 1999)

Mönnich, U. *Regular Description of Cross-Serial Dependencies* (MOL 6, Orlando, Florida, Jul. 1999)

Mönnich, U. und Morawietz, F. *Logical Description of Cross-Serial Dependencies* (FLoC '99, Workshop *Finite Model Theory and its Applications*, Trento, Jul. 1999)

Mönnich, U. und Morawietz, F. *Logical and Automata-Theoretic Approaches to Context-Free Tree Sets*, (Internationale Konferenz des SFB 340 *Linguistic Form and its Computation*, Bad Teinach, 10–13. Okt. 1999).

Mönnich, U. und Morawietz, F. *Monadische Prädikatenlogik zweiter Stufe in der Linguistik* (Kolloquium Computerlinguistik, CIS München, Feb. 2000)

Mönnich, U. und Morawietz, F. *Minimalismus in zwei regulären und logischen Schritten* (Oberseminar des Graduiertenkollegs *Sprache, Information, Logik*, CIS München, Feb. 2000)

Morawietz, F. *Chart Parsing as Constraint Programming* (IWPT 2000, Trento, Feb. 2000)

Morawietz, F. *MSO Logik und Linguistik* (Kurzseminar Semistrukturierte Daten, Universität Mainz, März 2000)

Mönnich, U. *Algebraic description of derivational minimalism* (AMILP 2000, Workshop *Algebraic Methods in Language Processing*, Iowa, Mai 2000)

Morawietz, F. *Derivational Minimalism in Two Regular and Logical Steps* (Tag+5, UFRL, TALaNa, Université Paris VII, Paris, Mai 2000)

Morawietz, F. *Chart Parsing and Constraint Propagation* (Coling 2000, Universität Saarbrücken, Aug. 2000)

Mönnich, U. und Morawietz, F. *A Formal Approach to Non-Context-Free Phenomena* (Kolloquium Allgemeine Linguistik, Universität Passau, Okt. 2000)

Mönnich, U. und Morawietz, F. *An Operational and Denotational Account of Second-Order Substitution* (Fakultätskolloquium Informatik, Universität Passau, Okt. 2000)

Mönnich, U. und Morawietz, F. *Attribuierter Minimalismus* (Potsdam-Tübinger Arbeitstreffen zur Daten-Problematik, Universität Tübingen, Okt. 2000)

Kühnberger, K.-U. *Anwendungen von Coalgebren in Linguistik und Philosophie* (Graduiertenkolleg *Algebraische, analytische und geometrische Methoden und ihre Wechselwirkung in der modernen Mathematik*, Bonn, Nov. 2000)

Kühnberger, K.-U. *Remarks on the Complexity of Truth* (Konferenz *Foundations of the Formal Sciences II*, Bonn, Nov. 2000)

Mönnich, U. und Morawietz, F. *A Two-Step Operational Description of TAGs* (Gemeinsamer Workshop University of Pennsylvania (IRCS) & Universitt Tbingen (SFB 441), Dez. 2000)

Einladungen/Gastwissenschaftler

Marcus Kracht *Syntax in Ketten* (20.07.1998)

Klaus Schulz *Entscheidbarkeit der Lösbarkeit von Kontextgleichungen mit zwei Variablen* (12.11.1998)

Hugo Volger *Principle Languages and Principle-based Parsing* (04.12.1998)

Workshop *Komplexitätstheoretische Eigenschaften von Grammatikformalismen* (18-19.06.1999)

Jens Michaelis *On varietes of minimalist requirements*
Christian Wartena *On the relation between LIGs and Multi-Pushdown automata*
Jochen Trommer *Formalizing Direct Optimaity Theory*
Marcus Kracht *Stacked Cases as a Substitute for Structure*
Peter Staudacher *Partial Wh-Movement and Compositionality*
Stephan Kepser *Linguistic Theories as Data Types*

Jens Michaelis *A Hierarchy of Minimalist Grammars* (13.09.1999)

Edward Stabler *Performance Models for a Derivational Minimalism* (10-14.10.1999)

Thomas Schwentick *Ausdrucksstarke und effiziente Abfragesprachen für baumartig strukturierte Daten* (25.11.1999)

Klaus Schulz *Aggregate als Antwortmodell bei semistrukturierten Daten* (11-13.04.2000)

Christian Wartena *Zur Komplexität dendraler Optimalitätssysteme* (11.05.2000)

Craig Thiersch *Head-Internal Relative Clauses?* (29.05.2000)

Christian Wartena *Datenstruktur und Satzstruktur – Anforderungen an die Satzanalyse* (27-28.10.2000)

Tom Cornell *Parsing and Grammar Engineering with Tree Automata* (27-28.10.2000)

Literatur

[1] Michel Bauderon and Bruno Courcelle. Graph expressions and graph rewriting. *Mathematical Systems Theory*, 20:83–127, 1987.

[2] Philippe Blache and Frank Morawietz. A non-generative constraint-based formalism. In Frank Morawietz, editor, *Some Aspects of Natural Language Processing and Constraint Programming*, number 150 in Arbeitspapiere des SFB 340, pages 1–28. Universität Tübingen, 2000.

[3] Michael Brody. *Lexico-Logical Form.* MIT Press, Cambridge, Ma., 1995.

[4] J. R. Büchi. Weak second-order arithmetic and finite automata. *Z. Math. Logik Grundlag. Math.*, 6:66–92, 1960.

[5] Noam Chomsky. Three models for the description of language. *I.R.E. Transactions of Information Theory*, 2:113–124, 1956.

[6] Noam Chomsky. On certain formal properties of grammars. *Information and Control*, 2(2):137–167, June 1959.

[7] Noam Chomsky. *Aspects of the Theory of Syntax.* MIT Press, Cambridge, Ma., 1965.

[8] Noam Chomsky. *The Logical Structure of Linguistic Theory.* Plenum, New York, 1975. (also 1985, UChicago Press).

[9] Noam Chomsky. *Lectures on Government and Binding.* Foris, Dordrecht, 1981.

[10] Noam Chomsky. *Knowledge of Language: Its Nature, Origin, and Use.* Praeger, New York, 1986.

[11] Noam Chomsky. *The Minimalist Program.* MIT Press, Cambridge, MA, 1995.

[12] Noam Chomsky. *The Minimalist Program*, volume 28 of *Current Studies in Linguistics.* MIT Press, 1995.

[13] Thomas L. Cornell. Island effects in type logical approaches to the minimalist program. In *Proceedings of FHCG-98 Joint Conference On Formal Grammar, Head-Driven Phrase Structure Grammar, and Categorial Grammar*, Saarbrücken, Germany, August 1998.

[14] Thomas L. Cornell. Derivational and representational views of minimalist syntactic calculi. In Alain Lecomte, François Lamarche, and Guy Perrier, editors, *Logical Aspects of Computational Linguistics (LACL '97)*, number 1582 in LNAI, pages 92–111. Springer, Berlin, 1999.

[15] Thomas L. Cornell. Parsing and grammar engineering with tree automata. In *Algebraic Methods in Language Processing: Second AMAST Workshop on Language Processing*, TWLT, University of Iowa, 2000. Parlevink.

[16] John Doner. Tree acceptors and some of their applications. *Journal of Computer and System Sciences*, 4:406–451, 1970.

[17] John E. Doner. Tree acceptors and some of their applications. *J. Comput. System Sci.*, 4:406–451, 1970.

[18] Heinz-Dieter Ebbinghaus and Jörg Flum. *Finite Model Theory.* Springer, Berlin, 1995.

[19] Calvin C. Elgot. Decision problems of finite automata design and related arithmetics. *Trans. Amer. Math. Soc.*, 98:21–51, 1961.

[20] Joseph E. Emonds. *A Unified Theory of Syntactic Categories.* Foris, Dordrecht, 1985.

[21] Joost Engelfriet and E.M. Schmidt. IO and OI, part I. *J. Comput. System Sci.*, 15:328–353, 1977.

[22] Joost Engelfriet and E.M. Schmidt. IO and OI, part II. *J. Comput. System Sci.*, 16:67–99, 1978.

[23] Ron Fagin. Generalized first-order spectra and polynomial-time recognizable sets. In R. Karp, editor, *Complexity of Computation*, volume 7 of *SIAM-AMS Proc.*, pages 27–41. AMS, 1974.

[24] Ron Fagin. Finite-model theory—a personal perspective. *TCS*, 116(1&2):3–31, 1993.

[25] Michael J. Fischer. Grammars with macro-like productions. In *Proceedings of the 9th Annual Symposium on Switching and Automata Theory*, pages 131–142. IEEE, 1968.

[26] Thom Frühwirth. Theory and practice of constraint handling rules. *Journal of Logic Programming*, 37(1–3):95–138, October 1998. Special Issue on Constraint Logic Programming.

[27] Sean Fulop and Edward L. Keenan. Compositionality: A global perspective. In U. Moennich and H. Kolb, editors, *Mathematics of Sentence Structure.* de Gruyter, Berlin, 1998. forthcoming.

[28] Henk Harkema. A recognizer for minimalist grammars. In *Sixth International Workshop on Parsing Technologies, IWPT'2000*, 2000.

[29] Neil Immerman. Languages that capture complexity classes. *SIAM J. of Computing*, 16(4):760–778, 1987.

[30] Edward L. Keenan. Further beyond the Frege boundary. In J. van der Does and J. van Eijck, editors, *Quantifiers, Logic, and Language*. CSLI Publications, Amsterdam, 1996.

[31] Edward L. Keenan. Logical objects. In M. Zeleny and C.A. Anderson, editors, *Alonzo Church Memorial Volume*. Kluwer, Boston, to appear.

[32] Edward L. Keenan and Edward P. Stabler. There is more than one language. In Léa Nash and Georges Tsoulas, editors, *Langues et Grammaire-1*, pages 217–235, Université Paris-8, 1994.

[33] Edward L. Keenan and Edward P. Stabler. Abstract syntax. In Anne-Marie DiSciullo, editor, *Configurations: Essays on Structure and Interpretation*, pages 329–344, Somerville, Massachusetts, 1996. Cascadilla Press.

[34] Edward L. Keenan and Edward P. Stabler. Syntactic invariants. In *6th Annual Conference on Language, Logic and Computation*, Stanford, 1997.

[35] Paul J. King. *A Logical Formalism for Head-Driven Phrase Structure Grammar*. PhD thesis, Manchester University, Manchester, England, 1989.

[36] Paul J. King. An expanded logical formalism for head–driven phrase stucture grammar. Arbeitspapiere des SFB 340 59, SFB 340, Universität Tübingen, 1994.

[37] Paul J. King. An expanded logical formalism for head-driven phrase structure grammar. Bericht 59, Arbeitspapiere des SFB 340, Tübingen/Stuttgart, 1994.

[38] Paul J. King, K. Simov, and B. Aldag. The complexity of modelability in finite and computable signatures of a constraint logic for head-driven phrase structure grammar. Universität Tübingen, Seminar für Sprachwissenschaft, 1997.

[39] Hans-Peter Kolb. Macros for minimalism? Towards weak descriptions of strong structures. In Hans-Peter Kolb and Uwe Mönnich, editors, *Mathematics of Syntactic Structure: Trees and Their Logics*, volume 44 of *Studies in Generative Grammar*. Mouton-de Gruyter, 1999.

[40] Hans-Peter Kolb, Jens Michaelis, Uwe Mönnich, and Frank Morawietz. An operational and denotational approach to non-context-freeness, 2000. Draft available under `http://tcl.sfs.nphil.uni-tuebingen.de/~frank/`.

[41] Hans-Peter Kolb, Uwe Mönnich, and Frank Morawietz. Descriptions of cross-serial dependencies. To appear in a special issue of *Grammars*, 2000. Draft available under `http://tcl.sfs.nphil.uni-tuebingen.de/~frank/`.

[42] F. W. Lawvere. *Functorial Semantics of Algebraic Theories*. PhD thesis, Columbia University, Braunschweig, 1963.

[43] J. Mezei and J. B. Wright. Algebraic automata and contextfree sets. *Information and Control*, 11:3–29, 1967.

[44] Jens Michaelis. Derivational minimalism is mildly context-sensitive. *Linguistics in Potsdam (LiP)* 5, Institut für Linguistik, Universität Potsdam, 1999. Available under `http://www.ling.uni-potsdam.de/~michael/`.

[45] Jens Michaelis. *On Formal Properties of Minimalist Grammars*. PhD thesis, University of Potsdam, 2001.

[46] Jens Michaelis, Uwe Mönnich, and Frank Morawietz. On minimalist attribute grammars and macro tree transducers, 2000. Draft available under `http://tcl.sfs.nphil.uni-tuebingen.de/~frank/`.

[47] Uwe Mönnich. Adjunction as substitution. In G-J. M. Kruijff, G.V. Morill, and R.T. Oehrle, editors, *Formal Grammar '97*, pages 169–178, Aix-en-Provence, 1997.

[48] Uwe Mönnich. Lexikalisch kontrollierte Kreuzabhängigkeiten und Kongruenzmorphologien. Talk at the Innovationskolleg Potsdam, 1997. Available under `http://tcl.sfs.nphil.uni-tuebingen.de/~tcl/uwe/uwe_moennich.html`.

[49] Uwe Mönnich. On cloning contextfreeness. In Hans-Peter Kolb and Uwe Mönnich, editors, *The Mathematics of Syntactic Structure*, number 44 in Studies in Generative Grammar, pages 195–229. Mouton de Gruyter, 1999.

[50] Frank Morawietz. Chart parsing and constraint programming. In *Proceedings of COLING-2000*, 2000.

[51] Frank Morawietz. Chart parsing as constraint propagation. In *Proceedings of the International Workshop on Parsing Technologies IWPT 2000*, Trento, 2000. ACL/SIGPARSE.

[52] Frank Morawietz. Chart parsing as constraint propagation. In Frank Morawietz, editor, *Some Aspects of Natural Language Processing and Constraint Programming*, number 150 in Arbeitspapiere des SFB 340, pages 29–50. Universität Tübingen, 2000.

[53] Frank Morawietz and Tomas L. Cornell. The Logic-Automaton Connection in Linguistics. In *Proceedings of LACL 1997*, number 1582 in LNAI. Springer, 1999.

[54] P. Stanley Peters and Robert W. Ritchie. On restricting the base component of transformational grammars. *Information and Control*, 18(5):483–501, June 1971.

[55] P. Stanley Peters and Robert W. Ritchie. On the generative power of transformational grammars. *Information Sciences*, 6:49–83, 1973.

[56] Carl Pollard and Ivan Sag. *Information-based Syntax and Semantics*. CSLI, Stanford, Ca, 1987.

[57] Carl Pollard and Ivan Sag. *Head-driven Phrase Structure Grammar.* University of Chicago Press, Chicago, 1994.

[58] Michael O. Rabin. Decidability of second order theories and automata on infinite trees. *Trans. Amer. Math. Soc.*, 141:1–35, 1969.

[59] Owen Rambow, K. Vijay-Shanker, and David Weir. D-Tree Grammars. In *Proceedings of ACL*, 1995.

[60] James Rogers. *Studies in the Logic of Trees with Applications to Grammar Formalisms.* PhD thesis, University of Delaware, 1994.

[61] John Robert Ross. *Constraints on Variables in Syntax.* PhD thesis, Massachusetts Institute of Technology, Cambridge, Ma, 1967.

[62] Stuart M. Shieber, Yves Schabes, and Fernando C. N. Pereira. Principles and implementation of deductive parsing. *Journal of Logic Programming*, 24(1–2):3–36, 1995.

[63] Edward Stabler. Derivational minimalism. In Christian Retoré, editor, *Logical Aspects of Computational Linguistics*, pages 68–95, Berlin, 1997. Springer. LNAI 1328.

[64] Edward Stabler. Minimalist grammars and recognition. Presented at the Workshop *Linguistic Form and its Computation* of the SFB 340 in Bad Teinach, 1999. Draft for the same volume.

[65] Edward Stabler. Remnant movement and complexity. In G. Bouma, G.-J. M. Kruijff, E. Hinrichs, and R. T. Oehrle, editors, *Constraints and Resources in Natural Language Syntax and Semantics*, volume II of *Studies in Constrained Based Lexicalism*, pages 299–326. CSLI, 1999.

[66] Edward P. Stabler and Edward L. Keenan. Structural similarity. In A. Nijholt, G. Scollo, T. Rus, and D. Heylen, editors, *Algebraic Methods in Language Processing, AMiLP 2000*, University of Iowa, 2000.

[67] J. W. Thatcher and J. B. Wright. Generalized finite automata theory with an application to a decision problem of second-order logic. *Math. Systems Theory*, 2(1):57–81, 1968.

[68] J. W. Thatcher and J. B. Wright. Generalized finite automata theory with an application to a decision problem of second-order logic. *Mathematical Systems Theory*, 2(1):57–81, 1968.

[69] W. Thomas. Languages, automata, and logic. In G. Rozenberg and A. Salomaa, editors, *Handbook of Formal Languages*, volume 3: Beyond Words, pages 389–455. Springer, Berlin, 1997.