

1995 IRCS TECHNICAL REPORT ABSTRACTS

Centering: A Framework for Modelling the Local Coherence of Discourse

Barbara J. Grosz
Aravind K. Joshi
Scott Weinstein
IRCS-95-01

This paper concerns relationships among focus of attention, choices of referring expression, and perceived coherence of utterances within a discourse segment. It presents a framework and initial theory of centering which are intended to model the local component of attentional state. The paper examines interactions between local coherence and choice of referring expressions; it argues that differences in coherence correspond in part to the inference demands made by different types of referring expressions given a particular attentional state. It demonstrates that the attentional state properties modelled by centering can account for these differences.

A First-Order Axiomatization of the Theory of Finite Trees

Rolf Backofen
James Rogers
K. Vijay-Shanker
IRCS-95-02

We provide first-order axioms for the theories of finite trees with bounded branching and finite trees with arbitrary (finite) branching. The signature is chosen to express, in a natural way, those properties of trees most relevant to linguistic theories. These axioms provide a foundation for results in linguistics that are based on reasoning formally about such properties. We include some observations on the expressive power of these theories relative to traditional language complexity classes.

A Lexicalized Tree Adjoining Grammar for English The XTAG Research Group

IRCS-95-03

This document describes a sizable grammar of English written in the TAG formalism and implemented for use with the XTAG system. This report and the grammar described herein supersedes the TAG grammar described in [Abeille et al., 1990]. The English grammar described in this report is based on the TAG formalism developed in [Joshi et al., 1975], which has been extended to include lexicalization ([Schabes et al., 1988]), and unification-based feature structures ([Vijay-Shanker and Joshi, 1991]). The grammar discussed in this report extends the grammar presented in [Abeille et al., 1990] in at least two ways. First, this grammar has more detailed linguistic analyses, and second, the grammar presented in this paper is fully implemented. The range of syntactic phenomena that can be handled is large and includes auxiliaries (including inversion), copula, raising and small clause constructions, topicalization, relative clauses, infinitives, gerunds, passives, adjuncts, it-clefts, wh-clefts, PRO constructions, noun-noun modifications, extraposition, determiner phrases, genitives,

negation, noun-verb contractions, sentential adjuncts and imperatives. The XTAG grammar has been relatively stable since November 1993, although new analyses are still being added periodically.

Preservation Theorems in Finite Model Theory

Eric Rosen
Scott Weinstein
IRCS-95-04

We develop various aspects of a finite model theory. We establish the optimality of normal forms over the class of finite structures and demonstrate separations among descriptive complexity classes. We establish negative results concerning preservation theorems. We introduce a generalized notion of preservation theorem and establish some positive results concerning “generalized preservation theorems” for first-order definable classes of finite structures which are closed under extensions.

CLiFF Notes: Research in the Language, Information and Computation Laboratory of the University of Pennsylvania

Annual Report: 1994, No. 4
Matthew Stone
Libby Levison (eds.)
IRCS-95-05

This report takes its name from the Computational Linguistics Feedback Forum (CLiFF), an informal discussion group for students and faculty. However the scope of the research covered in this report is broader than the title might suggest: this is the yearly report of the LINC Lab, the Language, Information and Computation Laboratory of the University of Pennsylvania.

It may at first be hard to see the threads that bind together the work presented here, work by faculty, graduate students and postdocs in the Computer Science and Linguistics Departments, and the Institute for Research in Cognitive Science. It includes prototypical Natural Language fields such as: Combinatorial Categorical Grammars, Tree Adjoining Grammars, syntactic parsing and the syntax-semantics interface; but it extends to statistical methods, plan inference, instruction understanding, intonation, causal reasoning, free word order languages, geometric reasoning, medical informatics, connectionism, and language acquisition.

Naturally, this introduction cannot spell out all the connections between these abstracts: we invite you to explore them on your own. In fact, with this issue it's easier than ever to do so: this document is accessible on the “information superhighway”. Just call up <http://www.cis.upenn.edu/~cliff-group/94/cliffnotes.html>

In addition, you can find many of the papers referenced in the CLiFF Notes on the net. Most can be obtained by following links from the authors' abstracts in the web version of this

report.

The abstracts describe the researchers' many areas of investigation, explain their shared concerns, and present some interesting work in the Cognitive Science. We hope its new online format makes the CLiFF notes a more useful and interesting guide to Computational Linguistics activity at Penn.

Programming Constructs for Unstructured Data

Peter Buneman
Susan B. Davidson
Dan Suciu
IRCS-95-06

We investigate languages for querying and transforming unstructured data, by which we mean languages than can be used without knowledge of the structure (schema) of the database. Such data can be represented using labeled trees, as suggested by ACeDB (A C. elegans Database), a database system popular with biologists, and more recently in Tsimmis, a system developed at Stanford for heterogeneous data integration. The approach we take is to extend structural recursion to labeled trees. This poses some interesting problems: first, it is no longer "flat" structural recursion, so that the usual syntactic forms and optimizations for collection types such as lists bags and sets may not be relevant. Second, we shall want to examine the possibility that the values we are manipulating may be cyclic. It is common in ACeDB, and generally in object-oriented databases, for objects to refer to each other, allowing the possibility of arbitrarily "deep" queries. Of course, such cyclic structures are usually constructed through the use of a reference/pointer type; however query languages are insensitive to these object identities and perform automatic dereferencing. We therefore want to understand what programs are well defined when we are allowed to make unbounded searches in the database.

Probabilistic Matching of Brain Images

J.C. Gee
L. LeBriquer
C. Barillot
D.R. Haynor
IRCS-95-07

Image matching has emerged as an important area of investigation in medical image analysis. In particular, much attention has been focused on the atlas problem, in which a template representing the structural anatomy of the human brain is deformed to match anatomic brain images from a given individual. The problem is made difficult because there are important differences in both the gross and local morphology of the brain among normal individuals. We have formulated the image matching problem under a Bayesian framework. The Bayesian methodology facilitates a principled approach to the development of a matching model. Of special interest is its capacity to deal with uncertainty in the estimates, a potentially important but generally ignored aspect of the

solution. In the construction of a reference system for the human brain, the Bayesian approach is well suited to the task of modeling variation in morphology. Statistical information about morphological variability, accumulated over past samples, can be formally introduced into the problem formulation to guide the matching or normalization of future data sets.

Bayesian Approach to the Brain Image Matching Problem

J.C. Gee
L. LeBriquer
C. Barillot
D.R. Haynor
R. Bajcsy
IRCS-95-08

The application of image matching to the problem of localizing structural anatomy in images of the human brain forms the specific aim of our work. The interpretation of such images is a difficult task for human observers because of the many ways in which the identity of a given structure can be obscured. Our approach is based on the assumption that a common topology underlies the anatomy of normal individuals. To the degree that this assumption holds, the localization problem can be solved by determining the mapping from the anatomy of a given individual to some referential atlas of cerebral anatomy. Previous such approaches have in many cases relied on a physical interpretation of this mapping. In this paper, we examine a more general Bayesian formulation of the image matching problem and demonstrate the approach on two-dimensional magnetic resonance images.

XTAG System - A Wide Coverage Grammar for English

Christy Doran
Dania Egedi
Beth Ann Hockey
B. Srinivas
Martin Zaidel
IRCS-95-09

This paper presents the XTAG system, a grammar development tool based on the Tree Adjoining Grammar (TAG) formalism that includes a wide-coverage syntactic grammar for English. The various components of the system are discussed and preliminary evaluation results from the parsing of various corpora are given. Results from the comparison of XTAG against the IBM statistical parser and the Alvey Natural Language Tool parser are also given.

Disambiguation of Super Parts of Speech (or Supertags):

Almost Parsing
Aravind K. Joshi
B. Srinivas
IRCS-95-10

In a lexicalized grammar formalism such as Lexicalized Tree-Adjoining Grammar (LTAG), each lexical item is associated

with at least one elementary structure (supertag) that localizes syntactic and semantic dependencies. Thus a parser for a lexicalized grammar must search a large set of supertags to choose the right ones to combine for the parse of the sentence. We present techniques for disambiguating supertags using local information such as lexical preference and local lexical dependencies. The similarity between LTAG and Dependency grammars is exploited in the dependency model of supertag disambiguation. The performance results for various models of supertag disambiguation such as unigram, trigram and dependency-based models are presented.

A Freely Available Syntactic Lexicon for English

Dania Egedi

Patrick Martin

IRCS-95-11

This paper presents a syntactic lexicon for English that was originally derived from the Oxford Advanced Learner's Dictionary and the Oxford Dictionary of Current Idiomatic English, and then modified and augmented by hand. There are more than 37,000 syntactic entries from all 8 parts of speech. An X-windows based tool is available for maintaining the lexicon and performing searches. C and Lisp hooks are also available so that the lexicon can be easily utilized by parsers and other programs.

Lexicalization and Grammar Development

B. Srinivas

Dania Egedi

Christy Doran

Tilman Becker

IRCS-95-12

In this paper we present a fully lexicalized grammar formalism as a particularly attractive framework for the specification of natural language grammars. We discuss in detail Feature-based, Lexicalized Tree Adjoining Grammars (FB-LTAGs), a representative of the class of lexicalized grammars. We illustrate the advantages of lexicalized grammars in various contexts of natural language processing, ranging from wide-coverage grammar development to parsing and machine translation. We also present a method for compact and efficient representation of lexicalized trees.

A Processing Model for Free Word Order Languages

Owen Rambow

Aravind K. Joshi

IRCS-95-13

Like many verb-final languages, German displays considerable word-order freedom: there is no syntactic constraint on the ordering of the nominal arguments of a verb, as long as the verb remains in final position. This effect is referred to as "scrambling", and is interpreted in transformational frameworks as leftward movement of the arguments. Furthermore, arguments from an embedded clause

may move out of their clause; this effect is referred to as "long-distance scrambling". While scrambling has recently received considerable attention in the syntactic literature, the status of long-distance scrambling has only rarely been addressed. The reason for this is the problematic status of the data: not only is long-distance scrambling highly dependent on pragmatic context, it also is strongly subject to degradation due to processing constraints. As in the case of center-embedding, it is not immediately clear whether to assume that observed unacceptability of highly complex sentences is due to grammatical restrictions, or whether we should assume that the competence grammar does not place any restrictions on scrambling (and that, therefore, all such sentences are in fact grammatical), and the unacceptability of some (or most) of the grammatically possible word orders is due to processing limitations. In this paper, we will argue for the second view by presenting a processing model for German.

On Descriptive Complexity, Language Complexity, and GB

James Rogers

IRCS-95-14

We introduce L2K,P, a monadic second-order language for reasoning about trees which characterizes the strongly Context-Free Languages in the sense that a set of finite trees is definable in L2K,P if it is (modulo a projection) a Local Set--the set of derivation trees generated by a CFG. This provides a flexible approach to establishing language-theoretic complexity results for formalisms that are based on systems of well-formedness constraints on trees. We demonstrate this technique by sketching two such results for Government and Binding Theory. First, we show that free-indexation, the mechanism assumed to mediate a variety of agreement and binding relationships in GB, is not definable in L2K,P and therefore not enforceable by CFGs. Second, we show how, in spite of this limitation, a reasonably complete GB account of English can be defined in L2K,P. Consequently, the language licensed by that account is strongly context-free. We illustrate some of the issues involved in establishing this result by looking at the definition, in L2K,P, of chains. The limitations of this definition provide some insight into the types of natural linguistic principles that correspond to higher levels of language complexity. We close with some speculation on the possible significance of these results for generative linguistics.

Some Novel Applications of Explanation-Based Learning to Parsing Lexicalized Tree-Adjoining Grammars

B. Srinivas

Aravind K. Joshi

IRCS-95-15

In this paper we present some novel applications of Explanation-Based Learning (EBL) technique to parsing Lexicalized Tree-Adjoining grammars. The novel aspects are (a) immediate generalization of parses in the training set, (b) generalizations over recursive structures and (c) representation of generalized parses as Finite State Transducers. A highly

impoverished parser called a “stapler” has also been introduced. We present experimental results using EBL for different corpora and architectures to show the effectiveness of our approach.

Reconstructing the Evolutionary History of Natural Languages

Tandy Warnow

Donald Ringe

Ann Taylor

IRCS-95-16

In this paper we present a new methodology for determining the evolutionary history of related languages. Our methodology uses linguistic information encoded as qualitative characters, so that prospective trees can be evaluated according to various optimization criteria, much as is done in the practice of inferring evolutionary history for biological species. By contrast with biology, however, we find that the linguistic data support evolutionary trees with extremely good compatibility scores, and that for such data it is possible to find optimal trees quickly. We have applied this method to the classification of Indo-European (IE) languages; we have been able to resolve one longstanding open problem (the Indo-Hittite hypothesis), and have indicated exactly what needs to be established in order to resolve another longstanding open problem (Italo-Celtic hypothesis). We have also discovered rather surprising facts about the history of Germanic within this family. Thus, this method provides an ability to resolve difficult questions in Historical Linguistics that have proved resistant to traditional character-based methodologies and to the more recent distance based approaches of lexicostatistics. The results of our methodology also indicate weaknesses in methods currently accepted and practiced in historical linguistics. One of our true cognates by traditional methods. Finally, this methodology permits the linguist to develop and test assumptions about the evolutionary relevance of different linguistic characters.

The Computational Analysis of the Syntax and Interpretation of “Free” Word Order in Turkish

(Ph.D. Dissertation)

Beryl Hoffman

IRCS-95-17

In this dissertation, I examine a language with “free” word order, specifically Turkish, in order to develop a formalism that can capture the syntax and the context-dependent interpretation of “free” word order within a computational framework. In “free” word order languages, word order is used to convey distinctions in meaning that are not captured by traditional truth-conditional semantics. The word order indicates the “information structure”, e.g. what is the “topic” and the “focus” of the sentence. The context-appropriate use of “free” word order is of considerable importance in developing practical applications in natural language interpretation, generation, and machine translation.

I develop a formalism called Multiset-CCG, an extension of Combinatory Categorical Grammars, CCGs, (Ades/Steedman 1982, Steedman 1985), and demonstrate its advantages in an implementation of a data-base query system that interprets Turkish questions and generates answers with contextually appropriate word orders. Multiset-CCG is a context-sensitive and polynomially parsable grammar that captures the formal and descriptive properties of “free” word order and restrictions on word order in simple and complex sentences (with discontinuous constituents and long distance dependencies). Multiset-CCG captures the context-dependent meaning of word order in Turkish by compositionally deriving the predicate-argument structure and the information structure of a sentence in parallel. The advantages of using such a formalism are that it is computationally attractive and that it provides a compositional and flexible surface structure that allows syntactic constituents to correspond to information structure constituents. A formalism that integrates information structure and syntax such as Multiset-CCG is essential to the computational tasks of interpreting and generating sentences with contextually appropriate word orders in “free” word order languages.

Parallel Programming Languages for Collections

(Ph.D. Dissertation)

Dan Suciu

IRCS-95-18

The thesis discusses the design, expressive power, and implementation of parallel programming languages for collections, the fragment dealing with collections of an object-oriented query language.

The Relational Algebra has a simple, intrinsic parallel semantics, which enabled the successful development of parallel relational database systems. But the implementation techniques of these systems do not carry over to the more complex object-oriented databases. In order to develop efficient parallel object-oriented database systems, one needs to (1) design their query languages with parallelism in mind, and (2) find new implementation techniques, specially designed for these languages. Here we pursue these goals for parallel languages for collections.

The collections of interest for us are sets, bags, and sequences (lists). We start by describing a basic collection calculus and additional forms of recursion on collections. They have an idealized parallel “execution”, assuming unbounded resources and instant communication, which gives us high-level parallel complexity measures.

An interesting fragment of the calculus expresses exactly the queries in the parallel complexity class NC. Here the salient construct is divide and conquer recursion on sets. Sublanguages obtained by imposing a bound k on the number of recursion nesting correspond to the subclasses AC^k , for $k \geq 1$.

We break the implementation of the calculus into three steps. First, sets and bags are implemented on sequences, using high-level parallel algorithms: we express such algorithms in a high-level language for sequences called MAP, built around a new form of recursion. Second, we describe a complexity-preserving compilation of MAP on a simple vector-parallel model. Third, we implement the vector model on a parallel multiprocessor. Here we choose as target the LogP model, which can be instantiated to simulate various multiprocessors. All but one of the vector model instructions require only restricted forms of communication patterns on LogP, called monotone communications. These in turn admit efficient implementations on LogP.

We ran two simple benchmarks on a LogP simulator, measuring the speedup and the scaleup. We report conditions under which good speedup and scaleup can be expected.

Prosodic Structure and Prosodically Constrained Syntax in Chinese

(Ph.D. Dissertation)

Shengli Feng
IRCS-95-19

An important motivation for exploring various linguistic phenomena is to gain deeper insight into the nature of human language. The interaction between prosody and syntax discussed in this dissertation may serve this purpose well. In Chinese, a classical syntactic problem is the following: only one constituent is, in general, allowed after the main verb. However, if object is a distressed element (a pronoun or a definite NP, for example), two constituents can legitimately occur after the verb. This pattern is explained by proposing a prosodic feature assignment on elementary trees in the Tree appearing in the Ba-construction. This is explained in terms of a weak prosodic non-branching verbal node at the end of a sentence.

It has been observed in modern Chinese that there is a tendency toward disyllabicity, and an opposite tendency toward monosyllabicity. These two contradictory tendencies are well-documented in the language. However, the literature lacks an explanation for why the two opposite tendencies should co-exist in the language. This problem is resolved by proposing two independent constraints operating at two different levels of grammar: The Monosyllabic Axiom (MA) applies at the morphemic level and the Foot Formation Rule (FFR) operates at the post-word or phrasal level. Hence, the output of MA is taken to be the input of FFR and these two constraints (MA and FFR) interact to determine the prosody of the language. It is argued that both MA and FFR contribute words to the lexicon; hence the primitive monosyllabic words and morphemes produced by the MA, the disyllabic (or trisyllabic) Basic-compounds derived from the notion of prosodic word (PrWd), and the four-syllable PrWd-compounds generated by a PrWd plus a PrWd are all stored in the dictionary.

It is suggested that the prosodic category of foot in Chinese originates from phonological change between Old Chinese (OC) and Middle Chinese (MC). It is seen that disyllabic foot formation has played and still plays a crucial role in the syntax and morphology of classical and modern Chinese. In Classical Chinese, prosody can be important in resolving questions of how word order changes take place. In both Classical and Modern Chinese, prosody is crucial to derive prosodic words and syntactic compounds. With the Edge-Theory proposed by McCarthy & Prince (1993) and the MA proposed in this thesis, an ALIGNMENT between syllables and morphemes can be derived as follows: ALIGN: [M]=[o]. Given this proposal, I argue that the lack of re-syllabification and bimoraic foot formation in Chinese can be derived from the Edge Alignment Constraint. Given the notion of Prosodic Word, compounds in Chinese can also be derived by an assignment of Foot Formation on idiomatic trees of a Lexicalized TAG. Compounds are therefore viewed as lexicalized Idiomatic PrWds.

The argument made here suggests that a full understanding of human language syntax and morphology must take prosody into account, and that the interaction between prosody and syntax is bi-directional: Syntax governs prosody and prosody also constrains syntax. Therefore, our model of grammar should be constructed in such a way that the prosodic structure is properly represented at the level of PF (Phonological Form), so that interaction can be accounted for within the grammar of a language and so that parametric variation among languages can be properly captured.

Tree Growth and Morphosyntactic Triggers in Adult SLA

Anne Vainikka

Martha Young-Scholten
IRCS-95-20

In previous papers (Vainikka & Young Scholten, 1994; in press a; and in press b), we proposed that in the acquisition of German, first language learners -- like second language learners -- gradually build up a syntactic structure. That is they posit only lexical projections at first, and then gradually posit the relevant functional projections.

In this paper, we examine the role of triggers in second language acquisition. Specifically, we ask: given that in second language, as in first language acquisition, syntactic projections gradually emerge, and given the notion that something 'becomes available' in the input to trigger the reorganization of the L2 grammar, exactly what triggers second language development?

We first describe the gradual building up of syntactic structures -- or tree growth -- in each of the stages of L2 acquisition (as described in Vainikka (1993/4)). We then consider what might constitute the relevant triggers of each of the stages of both L1 and L2 acquisitions (seen below).

We find that the status of triggers in first and second language

acquisition differs. We also observe that a number of learners in the ZISA studies (Clahsen & Muysken 1986) and in our LEXLEARN project in Dusseldorf appears to be fossilized. One might conclude that it is the different status of triggers for second language learners -- rather than lack of access to Universal Grammar -- that results in the lack of ultimate attainment of native competence. Since much of syntax is encoded in grammatical elements realized as affixes, difficulty in analyzing such affixes could seriously hamper language development.

What factors internal to the organism might be responsible for the difference between the treatment of triggers in L1 and L2 acquisition? Newport (1990) suggests that there may be a neurobiological factor relevant for the critical period which results in bound morphemes being processed by second language learners. We suspect, however, that ultimately the distinction between bound and free morphemes as triggers may be derived from phonology -- free morphemes typically constitute at least a phonological foot, while bound morphemes typically involve units smaller than a foot.

Proof Theoretic Approach to Specification Languages
Jawahar Lal Chirimar
(Ph.D. Dissertation)
IRCS-95-21

In this thesis I study FORUM as a specification language. FORUM is a higher-order logic based on the logical connectives of Linear Logic. As an initial example, I demonstrate that FORUM is well suited for specifying concurrent computations by specifying the higher-order (π) calculus. Next, I focus on the problem of specifying programming languages with higher-order functions, and imperative features such as assignable variables, exceptions and first-class continuations. I provide a modular and declarative specification of an untyped programming language, UML, which contains the above mentioned features. Further, I use the proof theory of FORUM to study program equivalence for the functional core of UML augmented with assignable variables. Using my compositional specifications in FORUM, I prove equivalence of programs that have been challenging for other specification languages. Finally I study the operation semantics of DLX, a prototypical RISC machine. I specify the sequential and pipelined operational semantics of DLX with important optimizations such as call-forwarding and early branch resolution, and prove them to be equivalent. Furthermore, I study the problem of code equivalence via the FORUM specification, and, in particular, analyze the problem of code rescheduling for DLX.

Verb Semantics for English-Chinese Translation
Martha Palmer
Zhibiao Wu
IRCS-95-22

A common practice in operational Machine Translation (MT)

and Natural language Processing (NLP) systems is to assume that a verb has a fixed number of senses and rely on a pre-compiled lexicon to achieve large coverage. This paper demonstrates that this assumption is too weak to cope with the similar problems of lexical divergences between languages and unexpected uses of words that give rise to cases outside of the pre-compiled lexicon coverage. We first examine the lexical divergences between English verbs and Chinese verbs. We then focus on a specific lexical selection problem -- translating English change-of-state verbs into Chinese verb compounds. We show that an accurate translation depends not only on information about the participants, but also on contextual information. Therefore, selectional restrictions on verb arguments lack the necessary power for accurate lexical selection. Second, we examine verb representation theories and practices in MT systems and show that under the fixed sense assumption, the existing representation schemes are not adequate for handling these lexical divergences and extending existing verb senses to unexpected uses. We then propose a method of verb representation based on conceptual lattices which allows the similarities among different verbs in different languages to be quantitatively measured. A prototype system UNICON implements this theory and performs more accurate MT lexical selection for our chosen set of verbs. An additional lexical module for UNICON is also provided that handles sense extension.

A Constructive Game Semantics for the Language of Linear Logic
Giorgi Japaridze
IRCS-95-23

I present a semantics for the language of first order additive-multiplicative linear logic, i.e. the language of classical first order logic with two sorts of disjunction and conjunction. The semantics allows to capture intuitions often associated with linear logic or constructivism such as sentences=games, sentences=resources or sentences=problems, where "truth" means existence of an effective winning (resource-using, problem solving) strategy.

The paper introduces a decidable first order logic ET in the above language and presents a proof of its soundness and completeness (in full language) with respect to this semantics. Allowing noneffective strategies in the latter is shown to lead to classical logic. Technically, my semantics is similar to Blass' game semantics (A. Blass, "A Game Semantics for Linear Logic," APAL, 56), and my completeness proof can likely be adapted to the logic corresponding to Blass' semantics to show its decidability. The reader needs to be familiar with classical (but not necessarily linear) logic and arithmetic.

Proving Properties of Typed Lambda-Terms Using Realizability, Covers, and Sheaves

Jean Gallier
IRCS-95-24

The main purpose of this paper is to take apart the reducibility method in order to understand how its pieces fit together, and in particular, to recast the conditions on candidates of reducibility as sheaf conditions. There has been a feeling among experts on this subject that it should be possible to present the reducibility method using more semantic means, and that a deeper understanding would then be gained. This paper gives mathematical substance to this feeling, by presenting a generalization of the reducibility method based on a semantic notion of realizability which uses the notion of a cover algebra (as in abstract sheaf theory). A key technical ingredient is the introduction of a new class of semantic structures equipped with preorders, called pre-applicative structures. These structures need not be extensional. In this framework, a general realizability theorem can be shown. Kleene's recursive realizability and a variant of Kreisel's modified realizability both fit into this framework. We are then able to prove a meta-theorem which shows that if a property of realizers satisfies some simple conditions, then it holds for the semantic interpretations of all terms. Applying this theorem to the special case of the term model, yields a general theorem for proving properties of typed lambda-terms, in particular, strong normalization and confluence. This approach clarifies the reducibility method by showing that the closure conditions on candidates of reducibility can be viewed as sheaf conditions. The above approach is applied to the simply-typed lambda-calculus (with types [figure], [figure], [figure], and [figure]), and to the second-order (polymorphic) lambda-calculus (with types [figure] and [figure]), for which it yields a new theorem.

Kripke Models and the (In)equational Logic of the Second-Order Lambda-Calculus

Jean Gallier
IRCS-95-25

We define a new class of Kripke structures for the second-order lambda-calculus, and investigate the soundness and completeness of some proof systems for proving inequalities (rewrite rules) as well as equations. The Kripke structures under consideration are equipped with preorders that correspond to an abstract form of reduction, and they are not necessarily extensional. A novelty of our approach is that we define these structures directly as functors $A:W \rightarrow \text{Preor}$ equipped with certain natural transformations corresponding to application and abstraction (where W is a preorder, the set of worlds, and Preor is the category of preorders). We make use of an explicit construction of the exponential of functors in the Cartesian-closed category Preor^w , and we also define a kind of exponential [equation unavailable in ASCII] to take care of type abstraction. However, we strive for simplicity, and we use only very elementary categorical concepts. Consequently, we believe that the models described in this paper are

more palatable than abstract categorical models which require much more sophisticated machinery (and are not models of rewrite rules anyway). We obtain soundness and completeness theorems that generalize some results of Mitchell and Moggi to the second-order lambda-calculus, and to sets of inequalities (rewrite rules).

Typing Untyped Lambda-Terms, or Reducibility Strikes Again!

Jean Gallier
IRCS 95-26

It was observed by Curry that when (untyped) lambda-terms can be assigned types, for example, simple types, these terms have nice properties (for example, they are strongly normalizing). Coppo, Dezani, and Veneri, introduced type systems using conjunctive types, and showed that several important classes of (untyped) terms can be characterized according to the shape of the types that can be assigned to these terms. For example, the strongly normalizable terms, the normalizable terms, and the terms having head-normal forms, can be characterized in some systems [figure] and [figure]. The proofs use variants of the method of reducibility. In this paper, we present a uniform approach for proving several meta-theorems relating properties of lambda-terms and their typability in the systems [figure] and [figure]. Our proofs use a new and more modular version of the reducibility method. As an application of our metatheorems, we show how the characterizations obtained by Coppo, Dezani, Veneri, and Pottinger, can be easily rederived. We also characterize the terms that have weak head-normal forms, which appears to be new. We conclude by stating a number of challenging open problems regarding possible generalizations of the realizability method.

Modal Logic over Finite Structures

Eric Rosen
IRCS-95-27

In this paper, we develop various aspects of the finite model theory of propositional modal logic. In particular, we show that certain results about the expressive power of modal logic over the class of all structures, due to van Benthem and his collaborators, remain true over the class of finite structures. We establish that a first-order definable class of finite models is closed under bisimulations if it is definable by a 'modal first-order sentence'. We show that a class of finite models that is defined by a modal sentence is closed under extensions if it is defined by a diamond-modal sentence. In sharp contrast, it is well known that many classical results for first-order logic, including various preservation theorems, fail for the class of finite models.

Finite Model Theory and Finite Variable Logics

Eric Rosen

(Ph.D. Dissertation)

IRCS-95-28

In this dissertation, I investigate some questions about the model theory of finite structures. One goal is to better understand the expressive power of various logical languages, including first-order logic (FO), over this class. A second, related, goal is to determine which results from classical model theory remain true when relativized to the class, F , of finite structures. As it is well-known that many such results become false, I also consider certain weakened generalizations of classical results.

I prove some basic results about the existential fragments of finite variable logics. I show that there are finite models whose existential k -variable theories are not finitely axiomatizable. I also establish the optimality of a normal form for infinitary existential k -variable logic, and separate certain fragments of this logic. I introduce a notion of a ‘generalized preservation theorem’, and establish certain partial positive results. I then show that existential preservation fails infinitary finite variable logic both over F and over the class of all structures. I also examine other preservation properties, e.g. for classes closed under homomorphisms.

In the final chapter, I investigate the finite model theory of propositional modal logic. I show that, in contrast to more expressive logics, modal logic is ‘well-behaved’ over F . In particular, I establish that various theorems that are true over the class of all structures also hold over F . I prove that, over F , a class of models is FO-definable and closed under bisimulations if it is defined by a modal FO sentence. In addition, I prove that, over F , a class is defined by a modal sentence and closed under extensions if it is defined by a diamond-modal sentence.

Attention in Early Scientific Psychology

Gary Hatfield

IRCS-95-29

Attention only “recently”—i.e. in the eighteenth century—achieved chapter status in psychology textbooks in which psychology is conceived as a natural science. This report first sets this entrance, by sketching the historical contexts in which psychology has been considered to be a natural science. It then traces the construction of phenomenological descriptions of attention, and compares selected theoretical and empirical developments in the study of attention over three time slices: mid eighteenth century, turn of the twentieth century, and late twentieth century. Significant descriptive, theoretical, and empirical continuity emerges when these developments are considered in the large. This continuity is open to several interpretations, including the view that attention research shows long-term convergence because it is conditioned by the basic structure of attention as a natural phenomenon, and the less optimistic view that theory making in at least this area of

psychology has been remarkably conservative when considered under large grain resolution, consisting in the reshuffling of a few core ideas.

First Order Logic, Fixed Point Logic and Linear Order

Anuj Dawar

Steven Lindell

Scott Weinstein

IRCS-95-30

Empty Subjects in Finnish and Hebrew

Anne Vainikka

Yonata Levy

IRCS-95-31

Unlike the traditionally discussed language types which either allow subject NP omission throughout (as do Italian and Chinese) or not at all (English or Swedish), the two languages described here are mixed languages. In Hebrew and Finnish, subject NPs can be omitted in certain persons or tenses, but not in others. In these languages omission of 1st and 2nd person subjects is common (as in Italian and Chinese), but in the 3rd person an overt subject NP is required (as in English and Swedish). This situation holds for all tenses in Finnish, and for tenses other than the present tense in Hebrew, where a subject NP is required in all persons.

The contribution of this paper is to provide an analysis of null subjects which both covers the complicated mixed systems of Hebrew and Finnish and extends to the systems of null subjects traditionally discussed in the literature. The analysis is based on the idea that the syntactic position of subject-verb agreement features varies cross-linguistically, and even within a language. Thus, in the English-type languages an overt subject is required to license the subject position, whereas in the Italian-type languages subject-verb agreement features occupy the subject position, and no overt subject NP is required. In Finnish and Hebrew, these features occur in the subject position in the 1st and 2nd person, but not in the 3rd person. In both languages, the agreement paradigm provides independent evidence for such an analysis, in that the 1st and 2nd person agreement suffixes resemble the corresponding pronouns, but the 3rd person suffixes do not bear such a resemblance.