3.1 Allgemeine Angaben zum Teilprojekt B7

3.11 Thema

Partial Parsing and the Acquisition of Lexical Syntax and Semantics

3.111 Fachgebiet und Arbeitsrichtung

Computational Linguistics, Parsing and Acquisition

3.12 Leiter

Steven Abney
Seminar für Sprachwissenschaft
Wilhelmstr. 113
Universität Tübingen
72074 Tübingen
Tel. 07071/29-4279

Mats Rooth
Seminar für Sprachwissenschaft
Wilhelmstr. 113
Universität Tübingen
72074 Tübingen
Tel. 07071/29-4279

3.13 Bisherige und beantragte Förderung des Teilprojekts (Ergänzungsausstattung)

Das Teilprojekt gehört zu den Neuanträgen; die Förderung fänge Januar 1995 an.

Haushaltsjahr	Personalmittel	Sächliche Verwaltungs-ausgaben	Investitionen	Gesamt
	100	101	102	103
bis 1991				
1992				
1993				
1994				
Zwischensumme				
1995	187,2	19,0	69,3	275,5
1996	187,2	19,0		206,2
1997	187,2	19,0		206,2

3.2 Zusammenfassung

We propose to investigate the logical problem of language acquisition, as applied to lexical representations. In particular, we assume that syntax and semantics are projected in a nontrivial sense from the lexicon. It is the lexical representations necessary to this projection of syntactic and semantic structure whose acquisition we wish to investigate. Moreover, we aim to forge a link between the linguistic and cognitive foundations of computational linguistics, and the statistical methods that are currently permitting computational linguistics to make great strides in lexical acquisition and broad-coverage parsing.

The acquisition of syntactic parameter values has received rather more attention than the acquisition of the lexicon. It is understandable that syntax has received more attention than the lexicon in studies of universal grammar, inasmuch as the lexicon is viewed in the first instance as the repository of idiosyncratic, unpredictable information. Most generalizations of interest are to be drawn in the area of syntax, or at best the interaction of syntax and lexical semantics. But the view that the lexicon is the repository of all that is idiosyncratic and unpredictable means that acquiring the lexicon is actually a much more challenging problem than acquiring the syntax. And as lexicalized grammars push more and more syntactic variability into the lexicon, the problem of acquiring lexical information becomes ever more challenging. That is the problem we would like to address with this project.

Because of the lexicon's variability and breadth, we expect that methods that are appropriate for its acquisition will not necessarily be traditional linguistic methods. Rather, novel corpus-based methods being developed in computational linguistics strike us as particularly well-suited to lexical acquisition. By looking at the behavior of words in large corpora, we intend to assign them to the appropriate lexical-semantic class provided by UG, and to determine more refined, word-specific behavior, particularly with respect to selectional restrictions.

Linguistically significant behavior is generally not observable in an unstructured stream of words. This introduces a bootstrapping problem. To determine their syntactic behavior, we must assign words to the correct syntactic and lexical-semantic categories, but to perform that assignment, we need to observe the words' syntactic behavior.

We address this problem by making two divisions. We divide the parsing problem into the subtask of grouping words into 'chunks', and the subtask of doing sentence-level parsing. And we divide the lexical information used in parsing into coarse category information, and fine-grained information for distinguishing the behavior of individual words. The fine-grained information consists predominantly in relations between pairs of words that are heads of their respective constituents, what we will call *head-head information.*

To 'prime the pump,' so to speak, we recognize chunks based on coarse category information. Given chunks, we can use coarse category information at the chunk level to do rough sentence-level parsing. This rough framework allows us, in turn, to induce finer-grained head-head information, particularly at the sentence level, but also within chunks (e.g., between nouns and the adjectives that modify them). This method can be iterated to improve the quality of acquired information. Induced head-head information can be used to improve chunk recognition and sentence-level parsing; and re-induction using improved skeletal parses yields cleaner head-head information.

Finally, it is important to note that lexical acquisition, as described here, represents a unique convergence of linguistic questions and more application-oriented computational issues. So often, one faces a choice between addressing scientific concerns and furthering practical ends. But here we have the happy situation that both kinds of considerations point in the same direction. So, for example, the division of parse trees into chunks and sentence-level dependencies also contributes to parser efficiency and robustness. And the emphasis on breadth of coverage is one of the most important criteria for practical applications.

3.3 Stand der Forschung

3.31 Projections

There is a growing consensus in the study of phrase structure within the GB framework that both functional elements and thematic elements participate fully in the $\overline{\mathrm{X}}$-system, in the sense that both types of element take subjects and complements, and project intermediate and maximal projections. At the same time, lexical elements are distinguished in that they are 'semantic' heads (s-heads) of larger-scale projections (s-projections) [1, 32, 65]. Thus, the modal is the syntactic head of the sentence, but verb is the s-head; likewise, determiner is the syntactic head of the noun phrase, but noun is the s-head.

The words that belong to a given s-projection often form a contiguous substring of the sentence. This is almost always the case in English—and more emphatically so if we recognize that prenominal adjectives and their adverb modifiers do not appear with functional satellites, hence, arguably, do not head independent s-projections. A contiguous set of words belonging to the same s-projection is a *chunk*. In the tree in figure 1, syntactic projections (*c-projections*) are represented by vertical lines connecting head to parent, s-projections are represented by circled sequences of nodes, and chunks are the boxed word-sequences.

An additional level of grouping, represented by the large dotted boxes in the figure, also proves useful. Define a *clause-projection* to be an s-projection that contains an IP node. A non-clausal s-projection belongs to the first clause-projection that dominates it. In most cases, the set of s-projections (chunks) belonging to the same clause-projection form a contiguous substring of the sentence. A contiguous sequence of chunks, all belonging to the same clause-projection, is a *simplex clause*. In figure 1, there are two clause-projections, one headed by *like*

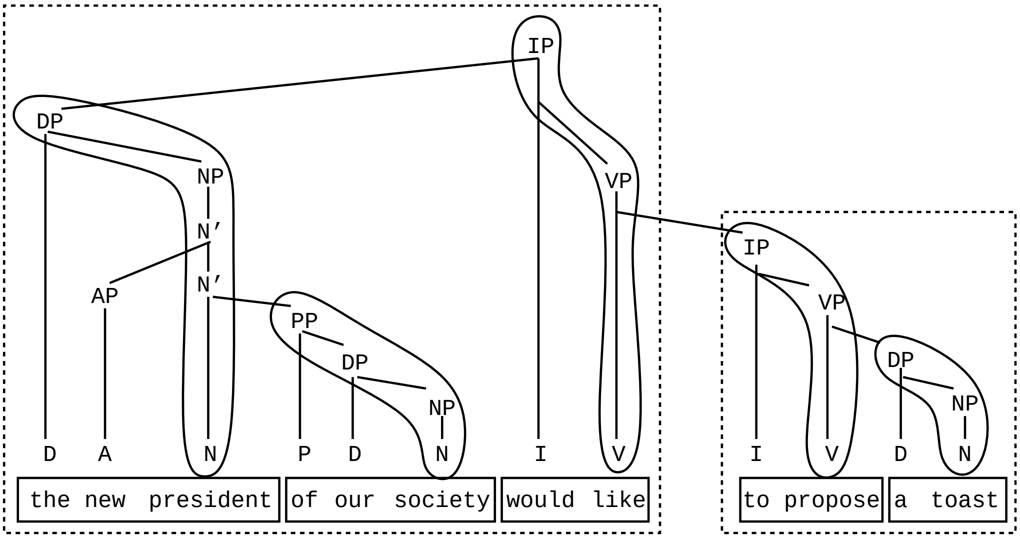

Abbildung 1: A tree partitioned into chunks. Circled groups of nodes are s-projections, chunks are boxed, and simplex clauses are marked by large dotted boxes.

and one headed by *propose*. Grouping together chunks that belong to the same clause-projection yields the simplex clauses marked by the large dotted boxes.

Chunks as phrasal units play a significant role in accounting for prosody, for certain kinds of psycholinguistic data, and for certain otherwise mysterious syntactic constraints [3, 7]. There is a series of psycholinguistic investigations into *performance structures*, that is, hierarchical sentence structures that are operationally defined. The measurements from which such structures emerge include naive parsing, transition error probabilities, and pausing [43, 48, 49, 23, 28, 33]. These structures have been noted to correspond closely to prosodic phrases [28, 59, 60], and have been used to generate prosody in text-to-speech systems [8]. What is of particular interest to us is their correspondence to chunks [7].

3.32 Parsing

Motivation of an entirely different sort for chunks and clauses comes from practical parsing considerations. In the last few years, statistical methods have been developed that have suddenly made it possible to deal with unrestricted natural text, instead of being constrained to "toy" domains [20, and numerous others]. Much of the effectiveness of these methods can be attributed to the possibility of modularizing the parsing problem, and attacking the pieces separately. So, for example, the first prominent success was obtained by isolating the problem of part-of-speech disambiguation. Disambiguating parts of speech proved to be soluble by a fairly simple statistical model [17, 22]. To proceed along these lines, it is natural to seek ways of breaking the remainder of the parsing problem into manageable pieces.

S-projections—or equivalently, chunks—provide an obvious possibility. We can divide the problem of assigning a structure to a sentence into three parts:

i. Identifying the s-projections, i.e., recognizing chunks;

ii. Assigning internal structure to s-projections; and

iii. Sentence-level parsing, i.e., assembling s-projections into a complete phrase structure.

The third task, sentence-level parsing, to a large extent consists in identifying head-head relations, for which we will also use the term *dependencies*.

Task (i), recognizing chunks, does turn out to be independently soluble, with respectable accuracy. Ejerhed [25, 24] discusses both finitary and stochastic methods for recognizing noun phrases and simplex clauses. Abney [2] describes a deterministic method for general chunks that achieves approximately 95% accuracy. And Rooth [57] describes a stochastic method for recovering noun phrase chunks with comparable accuracy. Numerous other researchers have used chunk-like units in applications where efficiency and robustness are at a premium, e.g., for data extraction [39], or induction of linguistic information from corpora [46]. Chunks have also been used to improve the robustness of traditional parsers [62].

Tasks (ii) and (iii) have not been addressed in the literature in the context of the parsing methodology sketched above. Our own pilot investigations indicate that the task of recognizing sentence-level dependencies can be solved with 90–95% accuracy by a small set of deterministic rules. The largest source of error is prepositional phrase attachment. Fortunately, statistical methods are known for effectively resolving PP attachment [36]. The task of assigning internal structure to s-projections has been previously addressed neither in the literature nor by us, but it is at least of limited scope, inasmuch as chunks rarely contain more than a handful of words.

We aim to improve current performance on each of these tasks by exploring a range of stochastic models and search strategies. We find it useful to classify parsing techniques along two dimensions. With respect to their approach to probabilities, we identify three types of parser:

- Non-stochastic, e.g. Abney's parser and Ejerhed's finitary parser;
- Bayesian ('reversing' a stochastic generation model), e.g. the standard optimization criteria for hidden Markov models and stochastic context free grammars;
- Regression models, including linear regression, CART [12], and transformation-based parsing [14].

The second dimension is search strategy. There are three categories of particular interest:

- Deterministic algorithms [47, 35, 37, 2];
- Exhaustive search, as in the Viterbi algorithm [64, 9, 54];
- Heuristic search, particularly best-first and beam search [51].

3.33 Lexical Representations

Representation of complementation/adjunction possibilities

Several recent theoretical approaches advocate restrictive definitions of the interface between syntax and the lexicon, where only certain lexical information is visible to the syntactic component [11, 53, 31]. A central idea is that the lexicon delivers a bag of complements, annotated with case and category information, and ordered by relative obliqueness (or deep order) of complements. (In certain theories, additional information is present, such as an optional marking of an external argument.) To a considerable extent, such representations are appropriate for the parsing model we have sketched. In particular, the representation is quite extensional, overtly representing the information about complementation possibilities that the parser requires. Also, for the most part, we can reasonably expect to acquire such representations, that is, to induce the correct representation for individual words based on their observable behavior. A few caveats are in order, however.

In theoretical approaches, a very restricted notion of complement is often employed, excluding statistically very important combinations. For instance, in some approaches the prepositional phrase in the combination *member of the board* is an adjunct rather than a complement. This association is nonetheless statistically important inasmuch as, in newspaper text, most occurrences of *member* are accompanied by a complement/adjunct *of*-phrase. Even in uncontroversial cases of adjunction, there seems to be a continuum of lexical conditioning, ranging from adjuncts which clearly introduce an external predication (such as *because*-clauses), through partially lexically conditioned adjuncts (such as spatial and temporal ones), to adjuncts which are quite intimately connected with the lexical semantics of the head. Once one undertakes the project of learning the combinatory properties of a large lexicon, one has to deal with a forbiddingly large variety of lexical conditioning.

The simplest move is to adopt a representation which does not distinguish between the two kinds of association. We think this is appropriate and perhaps unavoidable for a large class of lexically conditioned adjuncts. However, in other cases adjuncts are distinguished by their applicability to very broad classes of verbs, and adopting a distinct representation for adjunction is both theoretically and computationally attractive.

Selectional Constraints and Semantic Types

The term *selectional constraint* refers to semantic conditions imposed by verbs (more generally, heads) on the phrases they syntactically select. As these are typically conditions on the head of the selected phrase, they are a major source of constraint on the head-head dependencies that are the object of sentence-level parsing. Semantically, it is clear why such head-to-head relations give information about selection: usually, a single variable (or in other terminology, discourse referent) ends up as an argument of the head of the complement and of the verb. (Exceptions arise with intensional verbs and with intensional modifiers such as *fake* in a nominal argument.) For instance, transitive uses of aspectual verbs such as *begin* and *end* select event-denoting objects, such as *discussion* and *investigation*:

(1) a. ... the traffic-safety agency ended its investigation of the floor joints shortly after Thomas Built executives met with Deputy Transportation Secretary James Burnley .
b. He intends to begin an investigation of the matter.
c. Mr. Wright ended a discussion of the incident by saying , "Let's move on." ..
d. Walker Energy Partners said it began a preliminary discussion with a potential merger partner or partners , but it declined to name the suitor.

The semantic representation of *begin the discussion* includes the predications along the lines of $begin(x)$ and $discussion(x)$. Thus the nominal head *discussion* must at least be compatible with the property of being an event.

It is a commonplace in computational linguistics that selectional relations are required for filtering or ranking parses. As an example, consider the following.

(2) ... end the shareholder meeting on Friday.

Meeting could in principle be the head of the direct object of *end*, or it could be a post-modifier, with *shareholder* as the head of the direct object. The fact that *end* selects an event in the object position should be exploited to boost the ranking of the former analysis. The same point can be made about dependencies within chunks, such as dependencies between noun heads and their nominal, adjectival, and participial modifiers. In the example above, the fact that *shareholder* is a reasonable modifier for *meeting* supports an analysis where *meeting* is the head of the noun phrase chunk following *end.*

Selectional constraints (under many accounts) assume a hierarchy of semantic types: selection is not for specific lexical items, but for semantic types like *event* (in the example just given) or *human.* There is a large literature on semantic-type hierarchies, from many perspectives, including linguistic [41], artificial intelligence [26, 63], and psychological [50].

A hierarchy of semantic types useful in characterizing selection is likely to be finer-grained than what would be motivated by purely linguistic considerations. Generative work on lexical semantics has been concerned with identifying aspects of lexical representation which are relevant to syntax, such as the unergative/unaccusative distinction, deep complement order (or scale of obliqueness of complements), gross distinctions of semantic type of complements (individual vs. proposition vs. indirect question), gross aspectual distinctions (state vs. event/process), and the like. While some authors consider very fine classes of verb meaning [44], there is a consensus in recent work on a theory which captures linguistically relevant classes, while saying nothing about a wealth of distinctions which are semantically real, but linguistically inert [11, 31, 40]. In the more radical versions of these theories, agentive transitive process verbs with meanings as intuitively distinct as *read* and *eat* are treated as linguistically indistinguishable.

In a project concerned with bootstrapped extraction of lexical information from corpora, it is necessary to model the finer-grained semantic types, in addition to those which are relevant to the strictly linguistic module. For instance, we would want to take advantage of the obvious selectional differences between *read* and *eat* in a parser. Capturing fine-grained selectional distinctions would be desirable even if our primary goal were to induce linguistic lexical representations from examples of use. As has been pointed out in the literature on human acquisition of the lexicon, lexical sense ambiguity, syntactic ambiguity, and indeterminacy between complementation and adjunction make this task difficult. To take the simplest case, observing the sentences (3) below does not unambiguously indicate whether the language learner should induce a single lexical representation for *order* consistent with the semantics of both *pizza* and *retrial*, or two lexical representations.

(3) a. John ordered a pizza.
    b. The judge ordered a retrial.

In this case, the correct choice presumably involves two senses. A similar point arises with argument alternations; in the transitive-intransitive pair (4), the subject and object positions should be identified, while in (5), the two subject positions should be identified.

(4) a. The government decreased the tax rate.
    b. The price decreased.

(5) a. Her assistant wrote the tech report.
    b. My sister writes.

In both these cases, selectional constraints—of a kind which can be captured by a fine-grained semantic type model—would plausibly help in matching up the argument positions, and thus in identifying the correct linguistic lexical representations.

In computational linguistics, selectional constraints have been treated probabilistically [18]. Resnik's dissertation [55] combines a probabilistic account of selection preference with symbolic knowledge, the Wordnet psychological type hierarchy [50].

3.34 Acquisition

To summarize the discussion of the previous section, we require three types of lexical information:

i. The structured lexical representations that determine complementation and adjunction possibilities;

ii. An account of the semantic space, perhaps in the form of a semantic-type hierarchy; and

iii. Selectional constraints that relate the two, by putting semantic-type restrictions on the fillers of roles in the lexical representations of verbs.

A central concern of the proposed project is how these types of lexical information can be acquired. As linguists with a cognitive perspective, our ultimate concern is human acquisition, but we approach the problem at a certain level of abstraction. We intend to explore, in both breadth and detail, the question of how the information may in principle be acquired—what Lightfoot [45] calls the 'logical problem' of language acquisition.

If the class (i) of possible lexical-semantic representations is taken to be fixed, i.e., provided by universal grammar, the problem to be addressed is that of assigning individual words to classes. Here the literature on statistical classification problems is relevant. Diagnostics are known for assigning verbs to lexical-semantic classes, but many diagnostics used in the literature are accessible neither to a child nor to an automatic learner. On the other hand, we expect there to be surface properties that are informative with respect to class assignment, but that have not been identified in the literature, inasmuch as they only reveal themselves in statistical tendencies. Discovering such properties, and determining how to weigh evidence from different diagnostics, are problems with natural solutions in statistical classification techniques.

Selectional restrictions (iii) and semantic types (ii) can be induced from corpora. Church et. al. [19] pointed out that using a text corpus and a parser, one can get access to data relevant to selection by compiling tables of the frequency of head-head combinations: we can tell that *end* is an event-selecting verb by observing that it occurs frequently with complements the heads of which entail the property of being an event, for instance *discussion, trial, party* or *meeting.* The same frequency tables also contain information about the argument type hierarchy. We can recognize *discussion* as an event-denoting noun, for example, by observing that it occurs as the object of verbs like *begin, end*. Initial studies [13, 34, 36, 46, 55, 61] illustrate that both syntactic- and semantic-type frames can be induced by statistical techniques. One general procedure is to define a measure of distributional similarity, such as cosine or relative entropy [42], then to use the similarity measure to cluster words into classes [16, 27, 30, 52, 58]. Distributionally-defined classes typically do not correspond cleanly to semantic classes, in that there is a large confound of syntactic-class information in the distributional data. However, by using the chunk parser to select only pairs of content-word heads, we can control the syntactic environment, and in this way, we expect to eliminate most of the syntactic confound, and obtain much cleaner semantic classes.

3.4 Eigene Vorarbeiten

We have already mentioned most of our own relevant prior work. The first general account of projections of functional elements, and the concomitant distinction between c-projection (i.e. $\overline{\mathrm{X}}$-projection) and s-projection, is given in [1]. The relation between s-projections and chunks is proposed in [4], and further developed in [3, 7].

[2] describes a deterministic parser for chunks and simplex clauses, based on cascaded finite-state recognizers. The notion of pattern reliability (related to regression) is explored in [5, 6]. [57] describes a stochastic tagger and noun-chunk recognizer using Viterbi decoding.

In the area of selection, Hindle and Rooth [36] propose a probabilistic account of a syntactically restricted class of selectional restrictions, and show how they can be acquired from a large text corpus. The model is applied to disambiguating a class of prepositional phrase attachment ambiguities. [56] describes an experiment on semantic type induction based on EM estimation.

3.5 Ziele, Methoden, Arbeitsprogramm und Zeitplan

3.51 Ziele

Goals of the research are:

i. The construction of lexicalized, statistical, free text partial parsers for German and English.

ii. Investigation of the acquisition of lexical representations, with emphasis on semantic type information and representation of subcategorization/adjunction possibilities.

Subsidiary goals and products of the research are:

iii. Large automatically parsed corpora, for use in linguistic and lexicographic research.

iv. Computational lexicons induced from corpora.

v. Statistical language models for German and English, assigning probabilities to arbitrary sentences.

3.52 Methoden

Our general strategy is to take advantage of linguistic knowledge and hand-built rules when they are already available (as in the case of morphology and syntactic theories determining a motivated notion of chunk) or can be obtained with a moderate amount of effort (as in the case of chunk grammars), but to use automatic techniques to fill in much of the information used by the parser, in particular probability parameters. In order to allow different parsing strategies to employ common components, information and computational components are modularized; different sources of information are combined in a probabilistic framework. Finally, we aim to get preliminary versions of all essential components working quickly, in order to parse large corpora and extract information essential in later stages of research.

Below, we outline methods and computational approaches.

Chunk grammars and clause extraction

The chunk grammars will be obtained with conventional techniques, namely an iterative process of hand labeling modest amounts of data, writing grammars, and parsing corpora using a preliminary version of one of the parsers. For English, much of the relevant information can be extracted from treebanks. The grammars will be tuned and improved throughout the project, but chunk grammars for English and German with good coverage are to be written in an initial period of concentrated effort. This work is facilitated by the availability at SfS and IMS of German corpora labeled by hand with parts of speech (Stuttgart reference corpus), and of working part of speech taggers.

Reasonable performance in identification of clauses given chunks can be achieved with regular-expression matching [2, 24]. We will use this approach initially, subsequently investigating more systematic parsing techniques, and statistical approaches to clause identification.

Chunk parsing algorithms

We will develop several chunk parsing techniques in parallel, differing in search strategies and parameter estimation techniques. The initial phase of research aims at porting and extending chunk parsers using deterministic and Viterbi decoding techniques, and applying them to German and English corpus data. In later phases, we will incorporate selectional constraints into the parsers, and investigate other parsing techniques, in particular local statistical decision rules.

Abney [2] employed a deterministic chunk-recognition method. A regular-expression grammar for chunks is written by hand, and the longest-match heuristic is used to discriminate among alternative parses. Using this technique, it is easy to get a parser working quickly with quite good accuracy, on the order of 95% measured in chunks correctly recognized. This is an appropriate technique for use in the initial writing of chunk grammars.

Our application of Viterbi decoding will assume separate probability models for chunk contents (sequences of parts of speech constituting a chunk) and for the sentence-level contexts in which chunks appear. Given content and context models, an efficient search algorithm finds a globally optimal parse. The model described by Rooth [57] used a simple content model, a list of possible part of speech patterns with an associated probability distribution, and a trigram model of sentence-level context. The latter specifies a probability distribution over sentence-level phrase categories conditioned on the two linearly preceding phrase categories. Below, we will call this model a hidden Markov model (HMM) for chunks.

Left-to-right recognition

Viterbi decoding is a special case of the Bayesian method, in which one defines a probability distribution over partly-hidden structures, and in which parsing consists in reconstructing the most-probable complete structure that is consistent with the visible parts one is given as input. A disadvantage of this method is that it requires a global optimization; it is not possible to compute the probability of a particular piece of hidden structure during a left-to-right parse without estimating the probabilities of all relevant complete structures, which is expensive to do with any accuracy, and moreover implausible as a model of human performance.

For this reason, deterministic methods are of computational and psychological interest. The goal is to construct a model for the probability of a recognized chunk being correct. Probabilities can be conditioned on the state of the recognition automaton, as well as on cross-alternative properties such as the longest-match heuristic. This technique allows one to quantify the degree of uncertainty of local ambiguities, for example. In the limit, as local uncertainties approach zero, the left-to-right method approaches deterministic parsing.

Estimation of probability parameters

Large treebanks are available for English, and probability parameters can in many cases be computed directly as means in a training corpus. In the absence of a treebank, the training material consists simply of a text corpus, and probability

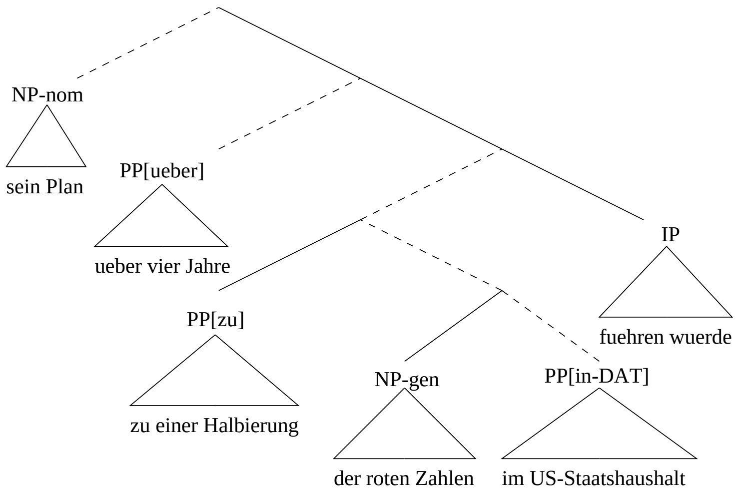

Abbildung 2: Dependencies in a German clause. Starting at a node, the licensing element can be identified by following the solid edges downwards to a labeled constituent. The licensed element is identified following one dotted edge, and then solid edges, downwards to a labeled constituent.

parameters must be derived indirectly. In a maximum likelihood framework, one aims to identify the parameters which maximize the probability of the observed corpus; this problem is computationally intractable in the cases of interest. However, constrained optimization techniques can improve the fit between model and data (specifically, the probability of the data given the model) up to a local maximum. We will emphasize the approach originated by Baum and his colleagues [10]. The *forward-backward* algorithm is used to iteratively re-estimate the parameters of hidden Markov models [54]. It can be viewed as a graph computation operating on a restricted kind of directed acyclic graph. The chunk HMM requires an extension to a general DAG, which we will implement.

Combined with the hand-written chunk grammar, we expect this approach to generate a statistical chunk parser for German comparable to what could be obtained with a parsed training corpus. At least in initial stages, we do not aim to induce chunk grammars themselves from data; rather, given chunk grammars, probability parameters of the chunk parser will be induced. We believe this is the quickest and surest route to a basic level of performance.

Recognizing and acquiring syntactic dependencies

The venue for this problem, which is of central importance and interest, is the area of syntax between the chunk and clause levels. Using the above methods for chunk and clause identification, a large corpus is mapped to a series of clauses, each partitioned into chunks. In order to identify structure within clauses, large

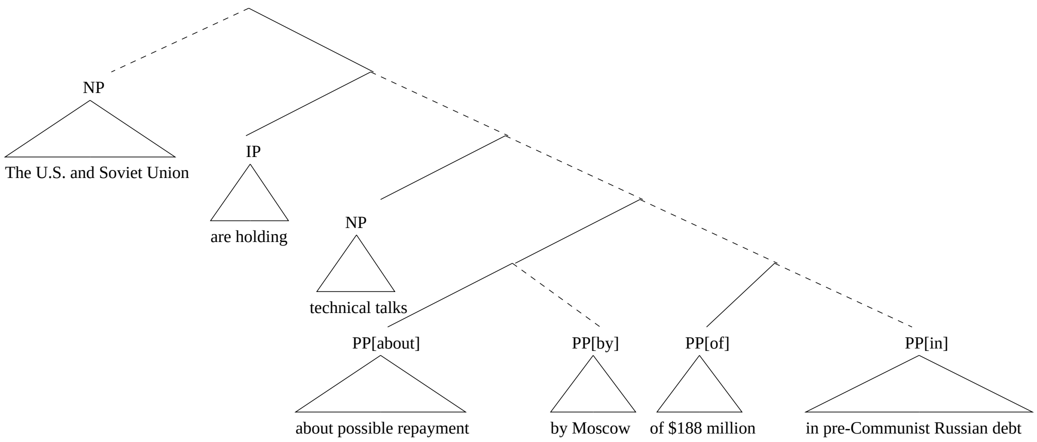

Abbildung 3: Dependencies in an English clause.

amounts of lexical information are required; we hypothesize that the best way to obtain this information is with a bootstrapping procedure. Initially, we will look at the problem of learning probabilistic subcategorization/adjunction frames annotated with only syntactic information, in particular category and case. This is to be done for a large lexicon, essentially the whole lexicon represented in a corpus of tens to hundreds of millions of words. A special case of this problem, restricted to certain simple grammatical relations, was studied for a large corpus and lexicon by Hindle and Rooth [36]. Other experiments have involved a broader class of frames, but have been limited to small samples of lexical items, small corpora, or have been based on statistical techniques which identify only the more common frames. Thus a broad-scale experiment has not yet been performed. Furthermore, we suspect the problem to be interestingly different in German, because of looser phrase order.

Two methods will be investigated, distinguished by the kinds of representations constructed. Imagine that the clauses in (6) have been identified and chunked by an an initial version of the parser.

(6) a. (daß) [sein Plan]$_{\text{DP-nom}}$ [über vier Jahre]$_{\text{PP[über]}}$ [zu einer Halbierung]$_{\text{PP[zu]}}$ [der roten Zahlen]$_{\text{DP-gen}}$ [im US-Staatshaushalt]$_{\text{PP[in-DAT]}}$ [führen würde]

b. [The U.S. and Soviet Union]$_{\text{DP}}$ [are holding]$_{\text{IP}}$ [technical talks]$_{\text{DP}}$ [about possible repayment]$_{\text{PP[about]}}$ [by Moscow]$_{\text{PP[by]}}$ [of $188 million]$_{\text{PP[of]}}$ [in pre-Communist Russian debts]$_{\text{PP[in]}}$

A dependency analysis of such a clause can be equated with a headed binary tree; the correct analysis of (6a) is the one in figure 2, and the correct analysis of (6b) is the one in figure 3. Each node represents a dependency between the s-head of its strong branch and the syntactic head of its weak branch; thus the top node

	asset	average	bit	bond	cent	cost	debt	dividend	foot	interest	mark	pence	point	price	rate	rating	security	share	stake	stock	tax	unit	value	yen
acquire	16		1			1	2			35							5	77	87	29		19		
boost		1		1		2	1	18		1	1		1	39	21	5		28	59	10	1		19	1
buy	16	2	2	48			2			53	1						36	348	107	190		29		2
climb					8	1					2	10	13	1										
cut						104	10	11					1	66	64	11			5	2	30		5	
decline		2	3		18	2	1						13	9	16			1		3				
drop	1	1	1		19				2		1	2	30	9	6	5								7
dump	1			3													2	10		10				
fall		2	1		132				1		2	14	171	38	33					1				28
gain					20					3	2	11	62		9			25	4		25			28
hold	18			7	1				1	22				3			5	68	121	30		3	1	
increase	6	3			2	25	5	26		8		1	3	26	36	2	1	36	75	2	11		16	
jump		2			3						2	8	9											
lower						20	2			1				23	83	55			16	1	2		4	
plunge					3						2		14	4						2				
purchase	8			5			2		1	17							6	95	24	20		6		
push	1	2		2		1				1			3	44	20		1	4		16	1	1	2	1
raise						23	5	28		8	5			131	149	26		5	74		46		11	1
reduce	9	1			1	76	105	3		5			1	22	55	8		9	41	2	26		21	
retain	1							1		13								17	21	1		3	1	
rise		13	9		136	18			2	2	3	52	125	18	19			1		1		1	2	22
sell	114	2	1	40			6			72	2		1	12	8		48	243	144	149		104		2
slash						17	4	9						20	6				1	1	3		3	
trade	1	1		2	2	2								9			7	22	2	37		5		1

Tabelle 1: Frequency counts for 24 verbs and 24 object heads.

represents a dependency between the s-head of [führen würde], namely [führen], and the nominative DP [sein Plan]. It is reasonable to enforce constraints such as the verb complex being the head of the clause, and noun phrases and prepositional phrases taking complements only on the right. This leaves us with a number of possible syntactic analyses in each case; the simplest strategy is to enumerate the analyses (possibly using a chart representation to share structure) in each step of an iterative re-estimation of lexical parameters. An alternative is to consider a flat representation of dependency possibilities between chunks, without building a tree or enumerating multiple analyses. This would have the disadvantage of not enforcing constituency constraints, but the possible computational advantage of reducing the number of representations computed.

As examples of the information to be learned by this procedure, the German lexicon should represent a statistical version of the fact that *führen* can take a *zu* prepositional dependent and that *Halbierung* can take a genitive dependent. The English lexicon should represent the fact that *repayment* can take both a *by* prepositional dependent and an *of* prepositional dependent. At this level one would not deal with semantic selection, for instance the fact that an DP headed by *Zahlen* is a semantically appropriate genitive dependent of *Halbieriung.*

Statistical models of selection and semantic type

Our basic approach is to use the parser to identify head-head dependencies in a corpus, and then treat the problem of describing what dependencies are possible as a matter of modeling word-word bigrams. This brings our project into contact with work on word n-gram language models for speech recoginition [9, etc.]. We intend to experiment with a variety of techniques discussed in this literature, but consider approaches involving classes or clusters [15, 52, 56] to be of particular interest, because in the context of our work, they appear to have a

	asset	bond	interest	security	share	stake	stock	unit	average	bit	cent	foot	mark	pence	point	yen	cost	debt	dividend	price	rate	rating	tax	value
acquire	16		35	5	77	87	29	19		1							1	2						
buy	16	48	53	36	348	107	190	29	2	2			1			2		2						
dump	1	3		2	10		10																	
hold	18	7	22	5	68	121	30	3			1	1								3				1
purchase	8	5	17	6	95	24	20	6				1						2						
retain	1		13		17	21	1	3											1					1
sell	114	40	72	48	243	144	149	104	2	1			2		1	2		6		12	8			
trade	1	2		7	22	2	37	5	1		2					1	2			9				
climb											8		2	10	13		1			1				
decline					1		3		2	3	18				13		2	1		9	16			
drop	1								1	1	19	2	1	2	30	7				9	6	5		
fall							1		2	1	132	1	2	14	171	28				38	33			
gain			3		25	4					20		2	11	62	28					9		25	
jump									2		3		2	8	9									
rise			2		1		1	1	13	9	136	2	3	52	125	22	18			18	19			2
plunge							2				3		2		14					4				
boost		1	1		28	59	10		1				1		1	1	2	1	18	39	21	5	1	19
cut						5	2								1		104	10	11	66	64	11	30	5
increase	6		8	1	36	75	2		3		2			1	3		25	5	26	26	36	2	11	16
lower			1			16	1										20	2		23	83	55	2	4
push	1	2	1	1	4		16	1	2						3	1	1			44	20		1	2
raise			8		5	74							5			1	23	5	28	131	149	26	46	11
reduce	9		5		9	41	2		1		1				1		76	105	3	22	55	8	26	21
slash						1	1										17	4	9	20	6		3	3

Tabelle 2: The same counts in another order.

natural foundation in lexical semantics.

Consider a concrete example, taken from Rooth [56]. Table 1 gives frequency counts for verb-object pairs in a corpus of about six million words, for twenty-four selected verbs and twenty-four selected nouns. The missing entries correspond to frequencies of zero in this corpus. However, some of the combinations with zero counts—for instance *dump stake* and *climb yen*—are intuitively plausible. The idea is to fill in these zero entries by identifying three semantic types in the table, represented by the diagonal blocks in table 2, where rows and columns have been rearranged. The types roughly correspond to (i) a concept of ownership; (ii) dimensions such as dollars and stock market points; (iii) abstract objects such as prices which move along linear scales. The types are two-dimensional, classifying both verbs and nouns. They can be used to fill in null entries in the sample, since while *dump stake* has a zero count, there is good evidence that both *dump* and *stake* belong in the first type.

The structure suggested by the blocks is given a probabilistic form in a *latent class* model of word bigrams. A verb-noun pair is viewed as being selected by (i) probabilistically picking a semantic type, and (ii) given the semantic type, independently picking a noun and verb using probability distributions characteristic of that type. Such models can be derived automatically; Pereira et al. [52] do this using a clustering technique, and Rooth [56] uses the statistical EM algorithm, a method of estimating latent class models proposed by Goodman [29]. Factor analysis techniques, such as singular value decomposition [21, 58] are also relevant.

In many cases, the two-dimensional clusters are semantically motivated, in that they are tied together by entailments of predications; thus the predications underlying the frequency counts in the first block have a common entailment that the entity corresponding to the object is owned (by somebody at some time).

Since the modeling of selection has a central role in this project, work on this will begin immediately, and proceed with some independence of development of the parsers. With a preliminary version of the chunk parser, several grammatical relations (such as verb-object in English, or head noun-modifier in German and English) can be reliably identified, and bigram tables computed from a large corpus. While these data will be limited to a few grammatical relations, they are realistic in the sense that they involve a large vocabulary and data drawn from a large corpus. We believe that they reflect nearly the full complexity of the problem of inducing selectional constraints and semantic types. Techniques for estimating probability distributions on word bigrams and inducing semantic types will be developed and tested with these data, before being integrated into the parser.

Integrating selection information and parsing

The parser architecture will be such that probabilistic theories of selection can be cleanly integrated. The techniques and results from the previous two sections will be combined, the goal being to induce a complementation/adjunction lexicon annotated with statistical models of the semantic type of dependent elements, in addition to their syntax.

In the case of models of chunks, we will be working with linguistically motivated grammars, and dependencies can be read off parses. One simple model is to view the lexical content of the head of each modifier phrase as being probabilistically conditioned on the head it modifies.

In the case of dependencies between chunks, we will investigate both the obvious technique of constructing a representation of trees determining dependencies in a given clause, and an approximate approach which evaluates possible head-head dependencies without constructing a representation of tree structures.

Induction of linguistic lexical representations

It is an open question whether the results of the learning procedure just outlined will contain information representing (either directly or indirectly) all linguistically relevant features; we suspect not. Building on the results of the above experiment, we will investigate the problem of learning such features from corpus data. The solution may lie in a refined theory of semantic types, with semantic notions such as agency having an a priori status (fixed by UG) in identifying linguistically significant classes of verbs.

3.53 Arbeitsprogramm und Zeitplan

A summary of the project work items is given in figure 4. The work can be divided roughly into three phrases, for which we may use the rubrics grammar,

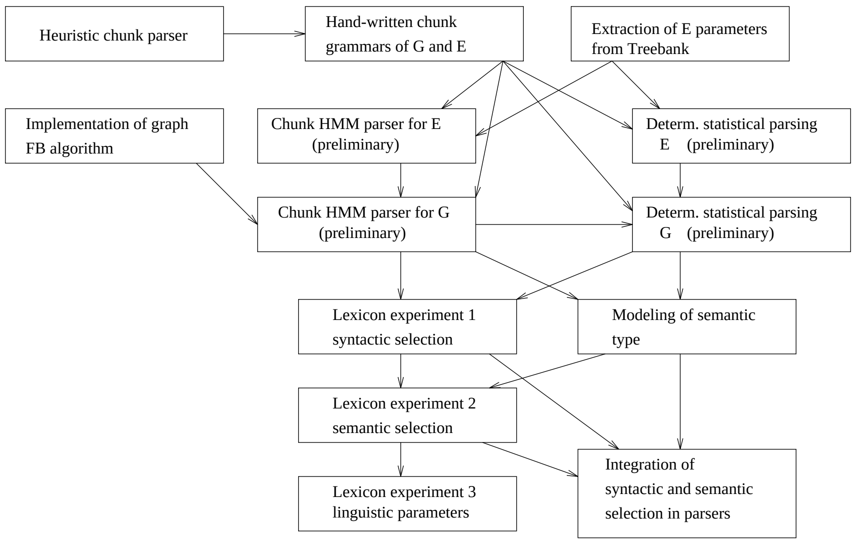

Abbildung 4: Dependencies between parts of the work plan.

parser, and acquisition. The grammar phrase (the top row in the figure) consists of constructing grammars for English and German with the aid of a quick-and-dirty chunk parser, as well as the extraction of statistical parameters for English from the UPenn Treebank. The parser phase (rows two and three) consists of constructing statistical chunk parsers for English and German. In English, we can use statistical parameters from the Treebank, but for German, we must first implement the graph forward-backward algorithm, for want of a parsed training corpus. In the acquisition phase (the bottommost three rows in the figure), we use the parsers to construct models of syntactic and semantic selection, feed those models back into the parsers, and tease apart the specifically linguistic from the more broadly cognitive components of the selection models.

The phases will not proceed in a strict sequence; we expect to continually refine all components of the complete system. Nonetheless, we do expect a concentration of activity on each phase in turn. We intend the grammar phase to go quickly, being the primary focus only for the first six months of the project. The aim is not a complete grammar of English and German, but only an adequate grammar for chunks and simplex clauses. The parser phase will take somewhat longer. Our goal is to complete the parser phrase, and turn our full attention to the acquisition phase, not later than the mid-point of the project period.

3.6 Stellung innerhalb des Programms des Sonderforschungsbereichs

The proposed project focuses, as do the other projects, on the syntax-semantics interface—in our case, the interaction between projection (cf. project A3 (Reis)) and lexical semantics (cf. projects A1 (Haider/Bierwisch), C3 (Kamp/Reyle)). The unique contribution of our project is the bridge it provides between linguistic foundations, particularly linguistically sophisticated lexical representations, and an exciting and explosively growing segment of computational linguistics, namely, work on statistical and corpus-based methods. The new statistical methods have suddenly permitted us to handle unrestricted text, with particular emphasis on acquisition and disambiguation, areas which remained stubbornly recalcitrant to traditional methods. The shortcoming of the new methods is their lack of linguistic depth, their missing linguistic foundations. We aim to synthesize the complementary strengths of sophisticated linguistic representations and statistical acquisition and disambiguation methods. We propose to show how linguistic representations of the kind developed in the lexical syntax/semantics projects support parsing of realistically large-scale natural text. Complementarily, we propose to show how the use of statistical methods supports exploratory data-driven linguistic study, allows us to determine the validity and scope of applicability of theoretical generalizations, and allows the results of the lexical syntax/semantics projects to be extended to a very broad empirical range.

The in-breadth methodology is of particular relevance to the fragment projects (B8, B9). The fragment projects also attempt to provide a sense of the big picture: the scope of the problem, and the interactions that arise among pieces. The acquisition component of our project aims to 'rough in' a fragment over a broad empirical domain. In a sense, the fragment projects aim for a consistent formalization and vertical integration of the lexical representations with other aspects of sentence grammar. We require coordination with the fragment projects, so as to use and acquire representations that are consistent with theirs. We are customers of the fragment projects in the sense that we intend to use their formalization of lexical representations, and hope to build on their grammatical descriptions at the chunk level. We also aim to provide the fragment projects with extensions of their lexica to broad domains, and probabilistic disambiguation techniques based on that lexical information.

The disambiguation component of our project aims to address a question of great importance for practical implementation: how do we choose the humanly-preferred analysis among the thousands of technically valid ones. Project C3 (Kamp/Reyle) attempts to limit the problem through the use of underspecified representations, and project B3 (Rohrer) develops inferential techniques for resolving lexical ambiguities. We complement this work, by addressing syntactic disambiguation as well as sense disambiguation, and by pursuing broad-coverage techniques.

The statistical disambiguation techniques we pursue also make a major contribution to parser efficiency—efficiency is indeed a *sine qua non* of corpus parsing. Probabilistic resolution of *local* ambiguities, our goal in the development of the left-to-right parsers, greatly enhances efficiency by eliminating most search. Particularly given our emphasis on the integration of probabilistic models with the linguistic representations of the syntax fragment project B8 (Hinrichs), we hope that the probabilistic techniques we develop will complement and be integrable with the techniques developed in project B4 (Hinrichs/Gerdemann) for efficient parsing of the formalism used in the fragment.

Finally, the relationship that we establish between syntactic projection (specifically, s-projection) and prosodic phrases provides a strong connection with the two projects that are concerned with the role of prosody in mediating between syntax and semantics: projects A4 (Drubig) and C4 (Dogil). Given the hypothesis that considerations from psycholinguistics, computational linguistics, prosody, and syntax converge on a unitary phenomenon, work on the phonetics, phonology, syntax and semantics of phrasing and related phenomena such as focus are relevant to the problem of defining a linguistically motivated notion of chunk. The provision of a robust parser for German is an additional link with project C4. Previous work on automatic accent assignment has used noun phrase identification as an important source of information [38]. And the chunk parser we propose to develop can be put to use in first-pass labeling of corpora with prosodic phrases; editing automated labelings has proven in previous annotation projects to be much more efficient than labeling from scratch.

3.7 Ergänzungsausstattung für das Teilprojekt

Es bedeuten:

PK: Personalbedarf und –kosten (Begründung vgl. 3.71)

SV: Sächliche Verwaltungsausgaben (Begründung vgl. 3.72)

I: Investitionen (Begründung vgl. 3.73)

PK	Bewilligung 1994		1995			1996			1997		
	Vergü-tungs-gruppe	An-zahl	Vergü-tungs-gruppe	An-zahl	Betrag in DM	Vergü-tungs-gruppe	An-zahl	Betrag in DM	Vergü-tungs-gruppe	An-zahl	Betrag in DM
	104	105	105a	106	107	108	109	110	111	112	113
			BAT II a	2	187,2	BAT II a	2	187,2	BAT II a	2	187,2
	zusammen		zusammen	2	187,2	zusammen	2	187,2	zusammen	2	187,2

SV		1995		1996		1997	
		Bezeichnung	Betrag in DM	Bezeichnung	Betrag in DM	Bezeichnung	Betrag in DM
		114	115	116	117	118	119
	515	Geräte		Geräte		Geräte	
	522	Verbrauchsmittel	5,0	Verbrauchsmittel	5,0	Verbrauchsmittel	5,0
	527	Reisekosten	10,0	Reisekosten	10,0	Reisekosten	10,0
	512	Bücher, Zeitschr.	2,0	Bücher, Zeitschr.	2,0	Bücher, Zeitschr.	2,0
	531a	Druck		Druck		Druck	
	531b	Kopien	2,0	Kopien	2,0	Kopien	2,0
		zusammen	19,0	zusammen	19,0	zusammen	19,0

I		1995	1996	1997
		Investitionen insgesamt DM	Investitionen insgesamt DM	Investitionen insgesamt DM
		120	121	122
		69,3		

3.71 Begründung des Personalbedarfs einschließlich Teilprojektleiter

	Name, akad. Grad, Dienststellung	engeres Fach des Mitarbeiters	Bezeichnung des Instituts der Hochschule bzw. der Einrichtung außerhalb der Hochschule	Anteil der aufgewendeten Gesamtzeit für das Teilprojekt in Std./Wo.	beratend	im SFB tätig seit	derzeitige Einstufung und beantragte Vergütungsgruppe
	123	124	125	129	130	131	132
Grundausstattung 3.71.1 *wissenschaftliche* Mitarbeiter (einschl. Hilfskräfte)	1. Abney, Stephen, Dr. 2. Rooth, Mats, Dr.	Comp.-Ling. Comp.-Ling.	SfS SfS	10 10			
3.71.2 *nichtwissenschaftliche* Mitarbeiter							
Ergänzungsausstattung 3.71.3 *wissenschaftliche* Mitarbeiter (einschl. Hilfskräfte)	3. N.N. 4. N.N.	Comp.-Ling. Comp.-Ling. Statistik		38.5 38.5			BAT IIa BAT IIa
3.71.4 *nichtwissenschaftliche* Mitarbeiter							

Responsibilities:

1. Project leader. Consultation in development of parsing and statistical algorithms, syntactic representations, leadership of project.

2. Project leader. Consultation in development of parsing and statistical algorithms, semantic representations, leadership of project.

3. Parsing algorithms, German chunk grammars, cognitive/linguistic models of parsing and the lexicon.

4. Statistical models and parameter estimation, word bigrams models, semantic type induction.

We require two BAT IIa positions, one for a specialist in stochastic models (4.) , and the second for a specialist in linguistic lexical representations and cognitive language processing (3.). Such researchers currently come from very different backgrounds. Specialists in stochastic models have backgrounds in computer science and related fields (including, oddly enough, physics: the mathematics of stochastic systems is the same whether the systems are physical or linguistic; we know of several examples of physicists who have done work on stochastic language models). Specialists in linguistic representations and cognitive processing issues generally come from linguistic or psycholinguistic backgrounds. It would be extremely difficult to find a single person with expertise in both fields.

3.72 Aufgliederung und Begründung der Sächlichen Verwaltungsausgaben (nach Haushaltsjahren)

	1995	1996	1997
Mittel für Neuanschaffung von Kleingeräten (515) sowie Verbrauchsmaterial (522)			
– Aus der *Grundausstattung* stehen vorraussichtlich zur Verfügung	6,5		
– Aus der *Ergänzungsausstattung* werden beantragt (vgl. Sp. 114–119)	5,0	5,0	5,0

Vorhandene Grundausstattung:

Two Macintosh IIX are available for the duration of the project, for use in preparation of publications and slides.

Verbrauchsmittel (522):

EDV-Zubehör, Büromaterial.

Reisekosten (527) Participation in scientific meetings. In 1995:

EACL

ACL

NEMLAP other conference/workshop dedicated non-traditional parsing methodology

Bücher und Zeitschriften (512):

Project-relevant dissertations and workshop proceedings not found in available libraries.

Kopien (531b):

Copies of circulated pre-publication papers relevant to the project.

3.73 Investitionen

Bezeichnung des Gerätes	beantragt für das Haushaltsjahr		
	1995	1996	1997
136	137	138	139
2 SPARCstation 20/50 64 MB RAM, 1,5 GB HD	69,3		

Die Investitionen bestehen im Einzelnen aus:

	Beschreibung	*Menge*	*Einzelpreis*	*Summe*
1.	Sparcstation 20/50	2	20.505,10	41.010,20
2.	AUI Adapter Kabel	2	147,60	295,20
3.	16 MB Memory expansion	4	1.809,50	7.238,00
4.	Festplatte Fujitsu 1GB	2	1.650,00	3.300,00
5.	S-Plus, single-user	1	3.990,00	3.990,00
6.	LDC Datenlizenz	1,5	3.000,00	4.500,00
			Summe	60.333,40
			Mehrwertsteuer	9.050,01
			Gesamtsumme	69.383,41

The workstations will be used for corpus experiments and stochastic modeling, both of which are computationally intensive. S-Plus is statistical software, required for the stochastic models. The Linguistic Data Consortium (LDC) license covers corpora, treebank data, and lexica necessary for the project. It is computed at half the annual rate (1,5 rather than 3 years), the remainder to be covered from other sources.

Literatur

[1] Steven Abney. *The English Noun Phrase in its Sentential Aspect.* PhD thesis, MIT, Cambridge, MA, 1987.

[2] Steven Abney. Rapid incremental parsing with repair. In *Proceedings of the 6th New OED Conference: Electronic Text Research*, pages 1–9, Waterloo, Ontario, October 1990. University of Waterloo.

[3] Steven Abney. Syntactic affixation and performance structures. In D. Bouchard and K. Leffel, editors, *Views on Phrase Structure.* Kluwer Academic Publishers, 1990.

[4] Steven Abney. Parsing by chunks. In Robert Berwick, Steven Abney, and Carol Tenny, editors, *Principle-Based Parsing.* Kluwer Academic Publishers, 1991.

[5] Steven Abney. Measures and models. In *Proceedings, Arpa Human Language Technologies Workshop*, San Mateo, CA, 1993. Morgan Kaufmann.

[6] Steven Abney. Reliability. In *Abstracts, Deutsche Gesellschaft für Sprachwissenschaft*, 1993.

[7] Steven Abney. Chunks and dependencies: Bringing processing evidence to bear on syntax. In *Computational Linguistics and the Foundations of Linguistic Theory.* CSLI, To appear.

[8] Joan Bachenko and Elizabeth Fitzpatrick. A computational grammar of discourse-neutral prosodic phrasing in English. *Computational Linguistics*, 16(3):155–170, 1990.

[9] L.R. Bahl, F. Jelinek, and R.L. Mercer. A maximum likelihood approach to continuous speech recognition. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, PAMI-5:179–190, 1983.

[10] L.E. Baum and J.A. Egon. An inequality with applications to statistical estimation for probabilistic functions of a markov process and to a model for ecology. *Bull. Amer. Meterol. Soc.*, 73:360–363, 1967.

[11] Manfred Bierwisch. *On the grammar of local prepositions.* In *Syntax, Semantik, und Lexicon, Studia Grammatica XXIX*, 1987. Akademie-Verlag, Berlin.

[12] L. Breiman, J. Friedman, R. Olshen, and C. Stone. *Classification and Regression Trees.* Wadsworth and Brooks, 1984.

[13] Michael R. Brent. Automatic acquisition of subcategorization frames from untagged, free-text corpora. In *Proceedings of the 29th Annual Meeting of the Association for Computational Linguistics*, pages 209–214, 1991.

[14] Eric Brill. *Transformation-Based Learning.* PhD thesis, Univ. of Pennsylvania, 1993.

[15] P. Brown, V. Della Pietra, S. Della Pietra, and R. Mercer. Class-based n-gram models of natural language. IBM internal research report, IBM, Yorktown Heights, New York 10598, 1990.

[16] P. Brown, V. Della Pietra, P. deSouza, J. Lai, and R. Mercer. Class-based n-gram models of natural language. *Computational Linguistics*, 18(4):467–480, 1992.

[17] Kenneth W. Church. A stochastic parts program and noun phrase parser for unrestricted texts. In *Proceedings of the Second Conference on Applied Natural Language Processing*, Austin, Texas, 1988.

[18] Kenneth W. Church, William Gale, Patrick Hanks, and Donald Hindle. Parsing, word associations and typical predicate-argument relations. In *International Workshop on Parsing Technologies*, pages 389–98, 1989.

[19] Kenneth W. Church, William A. Gale, Patrick Hanks, and Donald Hindle. Using statistics in lexical analysis. In U. Zernik, editor, *Lexical Acquisition*, Hillsdale, N.J.: Lawrence Erlbaum Associates, 1991.

[20] Kenneth Church and Robert Mercer. Introduction to the special issue on computational linguistics using large corpora. *Computational Linguistics*, 19(1):1–24, 1993.

[21] Scott Deerwester, Susan T. Dumais, George W. Furnas, Thomas K. Landauer, and Richard Harshman. Indexing by latent semantic indexing. *Journal of the American Society for Information Science*, 41(6):391–407, 1990.

[22] S. DeRose. Grammatical category disambiguation by statistical optimization. *Computational Linguistics*, 14(1), 1988.

[23] Jean-Yves Dommergues and Francois Grosjean. Performance structures in the recall of sentences. *Memory & Cognition*, 9(5):478–486, 1981.

[24] Eva Ejerhed. Finding clauses in unrestricted text by finitary and stochastic methods. In *Proceedings of the 2nd Conference on Applied Natural Language Processing.*, Austin, Texas, 1988.

[25] Eva Ejerhed and Kenneth Church. Finite state parsing. In Fred Karlsson, editor, *Papers from the Seventh Scandinavian Conference of Linguistics*, pages 410–432, Hallituskatu 11–13, SF-00100 Helsinki 10, Finland, 1983. University of Helsinki, Department of General Linguistics.

[26] Scott Elliot Fahlman. *NETL: A System for Representing and Using Real-World Knowledge.* The MIT Press, Cambridge, MA, 1979.

[27] Steven Paul Finch. *Finding Structure in Language.* PhD thesis, University of Edinburgh, 1993.

[28] James Paul Gee and François Grosjean. Performance structures: A psycholinguistic and linguistic appraisal. *Cognitive Psychology*, 15:411–458, 1983.

[29] Goodman. The analysis of systems of qualitative variables when some of the variables are unobservable. *American Journal of Sociology*, 79:1179-1259, 1974.

[30] G. Grefenstette and M. Hearst. A knowledge-poor method for refining automatically-discovered lexical relations: Combining weak techniques for stronger results. In *AAAI Workshop on Statistically-Based NLP Techniques, Tenth National Conference on Artificial Intelligence*, July 1992.

[31] Jane Grimshaw. *Argument structure.* Cambridge, Mass.: MIT Press, 1990.

[32] Jane Grimshaw. Extended projections. Manuscript, Rutgers University, 1993.

[33] F. Grosjean, L. Grosjean, and H. Lane. The patterns of silence: Performance structures in sentence production. *Cognitive Psychology*, 11:58–81, 1979.

[34] D. Hindle. Noun classification from predicate-argument structures. In *Proceedings of the 28th Annual Meeting of the Assocation of Computational Linguistics, Pittsburgh, Penna.*, pages 268–275, 1990.

[35] Donald Hindle. User manual for Fidditch. Technical Memorandum #7590-142, Naval Research Laboratory, 1983.

[36] D. Hindle and M. Rooth. Structural ambiguity and lexical relations. *Computational Linguistics 18*, 1993. Shorter version in *Proceedings of DARPA Speech and Natural Language Workshop.* Morgan Kaufman: New York, June 1990.

[37] Donald Hindle. A parser for text corpora. In A. Zampolli, editor, *Computational Approaches to the Lexicon*, New York : Oxford University Press, 1994.

[38] Julia Hirschberg. Pitch accent in context: predicting intonational prominence from text. *Artificial Intelligence*, 63: 305-340.

[39] Jerry R. Hobbs, Douglas Appelt, Mabry Tyson, and Megumi Kameyama. Fastus: A system for extracting information from text. In *ARPA Workshop on Human Language Technology*, San Mateo, CA, 1993. Advanced Research Projects Agency (ARPA), Morgan Kaufmann.

[40] Ray S. Jackendoff, *Semantic structures.* Cambridge, Mass.: MIT Press, 1990.

[41] Jerrold J. Katz and Paul M. Postal. *An Integrated Theory of Linguistic Descriptions.* Research Monograph No. 26. The MIT Press, Cambridge, MA, 1964.

[42] S. Kullback and R. A. Leibler. On information and sufficiency. *Ann. Math. Stat.*, 22:79-86, 1951.

[43] W.J.M. Levelt. Hierarchial chunking in sentence processing. *Perception & Psychophysics*, 8(2):99–103, 1970.

[44] Beth Levin. *English verb classes and alternations : a preliminary investigation.* Chicago : University of Chicago Press, 1993.

[45] David Lightfoot. *The Language Lottery: Toward a Biology of Grammars.* The MIT Press, Cambridge, MA, 1982.

[46] Christopher D. Manning. Automatic acquisition of a large subcategorization dictionary from corpora. In *31st Annual Meeting of the Association for Computational Linguistics*, pages 235–242, 1993.

[47] Mitchell Marcus. *A Theory of Syntactic Recognition for Natural Language.* The MIT Press, Cambridge, MA, 1980.

[48] Edwin Martin. Toward an analysis of subjective phrase structure. *Psychological Bulletin*, 74(3):153–166, September 1970.

[49] James G. Martin. Rhythmic (hierarchical) versus serial structure in speech and other behavior. *Psychological Review*, 79(6):487–509, 1972.

[50] G.A. Miller, R. Beckwith, C. Fellbaum, D. Gross, and K.J. Miller. Introduction to Wordnet: An on-line lexical database. *Journal of Lexicography*, 3(4), 1990.

[51] Judea Pearl. *Heuristics: Intelligent Search Strategies for Computer Problem Solving.* Addison-Wesley Publishing Company, Reading, MA, 1984.

[52] Fernando Pereira, Naftali Tishby, and Lillian Lee. Distributional clustering of English words. In *31st Annual Meeting of the Association for Computational Linguistics*, pages 183–190, 1993.

[53] Carl Pollard and Ivan A. Sag. *Information-Based Syntax and Semantics.* CSLI Lecture Notes Number 13. CSLI, 1987.

[54] L. R. Rabiner. A tutorial on Hidden Markov Models and selected applications in speech recognition. *Proceedings of the IEEE*, 77(2):257–285, February 1989.

[55] Philip Resnik. *Selection and Information.* PhD thesis, University of Pennsylvania, Philadelphia, PA, 1993.

[56] Mats Rooth. Latent selection classes. Unpublished manuscript.

[57] Mats Rooth. Unitary stochastic part-of-speech and phrase tagging. Submitted to COLING 94.

[58] Hinrich Schuetze. Word space. In S.J. Hanson, J.D. Cowan, and C.L. Giles, editors, *Advances in Neural Information Processing Systems 5.* Morgan Kaufmann, San Mateo CA, 1993.

[59] Elisabeth O. Selkirk. The role of prosodic categories in English word stress. *Linguistic Inquiry*, 11(3):563–605, 1980.

[60] Elisabeth O. Selkirk. On the nature of phonological representations. In T. Myers, J. Laver, and J. Anderson, editors, *The Cognitive Representation of Speech.* North-Holland Publishing Company, Amsterdam, 1981.

[61] Frank Smadja. *Extracting Collocations from Text. An Application: Language Generation.* PhD thesis, Columbia University, New York, NY, 1991.

[62] David Stallard and Robert Bobrow. Fragment processing in the delphi system. ms., BBN, 1992.

[63] David S. Touretzky. *The mathematics of inheritance systems.* Pitman, London, 1986.

[64] A. Viterbi. Error bounds for convolution codes and an asymptotically optimum decoding algorithm *IEEE Trans. Inform. Theory*, vol IT-13, pp260-269, 1967.

[65] Joost Zwarts. *X′-Syntax, X′-Semantics.* PhD thesis, Utrecht University, 1992.