TILBURG UNIVERSITY

SHOE

Powers, D.M.W.; Daelemans, W.M.P.

Document version:
Publisher final version (usually the publisher pdf)

Publication date:
1991

Link to publication

Citation for published version (APA):
Powers, D. M. W., & Daelemans, W. M. P. (1991). SHOE: The extraction of hierarchical structure for machine learning of natural language. (ITK Research Memo). Tilburg: Institute for Language Technology and Artifical Intelligence, Tilburg University.

General rights
Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal

Take down policy
If you believe that this document breaches copyright, please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Download date: 04. jan. 2016

CBM
CBM
R
8419
1991
10

CINO0291

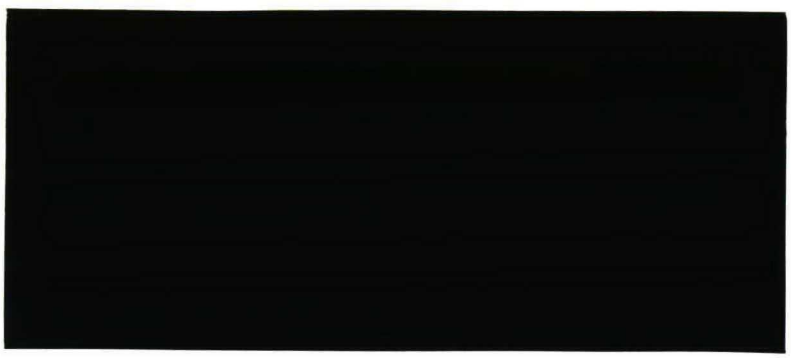
[redacted]

ITK

MEMO

[redacted]

INSTITUTE FOR LANGUAGE TECHNOLOGY AND ARTIFICIAL INTELLIGENCE INSTITUUT VOOR TAAL- EN KENNISTECHNOLOGIE



K.U.B.
BIBLIOTHEEK
TILBURG

ITK Research Memo
december 1991

SHOE: The Extraction of Hierarchical Structure for Machine Learning of Natural Language.

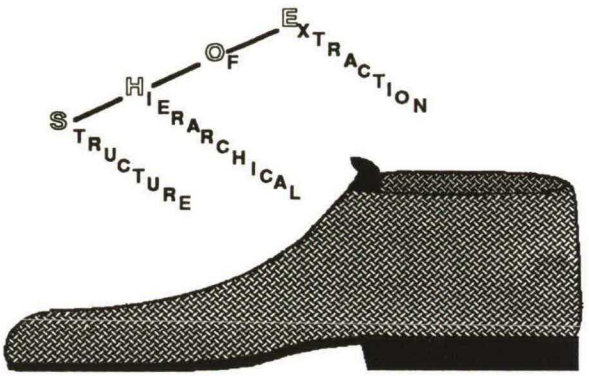
D. Powers & W. Daelemans

No. 10

©1991. Institute for Language Technology and Artificial Intelligence,
Tilburg University, P.O.Box 90153, 5000 LE Tilburg, The Netherlands
Phone: +3113 663113, Fax: +3113 663110.

Project Proposal
for the

ESPRIT
BASIC RESEARCH ACTION



SHOE

Extraction of Hierarchical Structure for Machine Learning of Natural Language

SHOE: The Extraction of Hierarchical Structure for Machine Learning of Natural Language. A project proposal *

David Powers & Walter Daelemans (eds.)

Abstract

The goal of the SHOE project is to force a breakthrough in the development of Machine Learning techniques for the extraction of hierarchically structured knowledge in a selection of linguistic domains, with a particular emphasis on extending and characterizing the limits and capabilities of unsupervised learning techniques.

We will explore different machine learning techniques focused on the lower levels of language and a bottom up recognition of structure in language, and are thus couching the project in terms of Extraction Of Hierarchical Structure (SHOE). Hierarchical structure is a property of knowledge in all areas of natural language processing. Abstraction (generalization) hierarchies of rules and representations are a salient feature of phonological, morphological, and lexical (semantic and syntactic) knowledge.

We will relate different Machine Learning techniques in a narrow subdomain, and will conversely seek to characterize the relative complexity of language subproblems by applying the same ML techniques across the different aspects of language, at adjacent hierarchical levels in each submodality, and between hierarchies and levels. This latter focus is fundamental to a proper semantics (understood in terms of interrelationships between the lexical and ontological hierarchies), but has been a glaring omission in most existing work.

From the perspective of Machine Learning, the primary focus has been on supervised learning techniques in which detailed information and/or interactive assistance is assumed to be available. Interhierarchical correspondences promise to provide such additional teacher/critic/supervisor information automatically.

We do not immediately want to tackle the higher level subtleties of language which most NL work is concerned with, but to work on other important problems such as dealing with uncertainty (speech recognition and optical character recognition), stress and syllable structure (phonology), acquisition of hierarchical lexicons for unification-based grammars and assimilation of new words, proper names etc. (lexicon), as well as the interaction of syntax, semantics and ontology in prepositional usage, subcategorization, etc. These areas have been selected with the knowledge acquisition bottlenecks of Esprit projects such as Acquilex, Sundial, and Plus in mind - problems that are representative for speech and language technology as a whole.

*This is an abridged version of a research project proposal submitted to Esprit Basic Research, October 14th, 1991. We especially want to acknowledge the contributions of Steven Gillis (UIA) to the text of the proposal.

Contents

1 Introduction 3

2 Aims of the Project 3

2.1 New learning techniques for natural language processing 3
2.2 A matrix of language and learning technologies 6
2.3 Interdisciplinary analysis of existing techniques 7
2.4 Technical and commercial spin-off 8
2.5 Summary of aims 9

3 Background of the Project 10

3.1 Learning methods and complexity 10
3.2 Interdisciplinary insights and cross-application 11
3.3 Representation, operationality and standardization 12
3.4 Ontology and grounding 12
3.5 Sensory-Motor 13

4 Description of the Project 14

4.1 Bootstrapping and negative information 14
4.2 Incremental Learning 17
4.3 Classification, clustering and taxonomy 18
4.4 Noise, errors, complexity and the teacher 19
4.5 Implicit and Explicit Teacher and Critic 20
4.6 Concepts and contexts, paradigm and metaphor 21
4.7 Relationships between learning methods 22
4.8 Complex Dynamics for Natural Language Processing 22
4.8.1 Language areas 23
4.8.2 Complex dynamics techniques 23
4.8.3 Self-organization of phonological systems 24
4.8.4 Evolution of innate syntactic architectures 25
4.8.5 Acquisition of lexical knowledge from corpora 25
4.8.6 Data level parallelism 26

5 References 28

1 Introduction

Our prime motivation for disseminating a project proposal in the present form is to encourage discussion about the field of Machine Learning of Natural Language (MLNL) in general and about our approach to the subject in particular. We believe that MLNL will become increasingly important in AI and NLP not only because of its potential to alleviate knowledge acquisition and adaptation bottlenecks in language technology, but also because of its theoretical interest both for (computational) linguistics and for machine learning.

The consortium submitting the proposal consists of the following (associate) partners and key researchers.

DFKI (German AI Institute). *Joerg Siekmann, David Powers, Franz Schmalhofer.*

ITK (Institute for Language Technology and AI). *Harry Bunt, Walter Daelemans, Peter Flach.*

University of Osnabrück. *Claus-Rainer Rollinger, Werner Emde.*

University of Trier. *Burghard Rieger, Jürgen Schrepp, Sven Naumann.*

Psychology Department, Tilburg University. *Beatrice De Gelder, Jean Vroomen.*

Computational Linguistics, University of Amsterdam. *Remko Scha, Willem Meijs, Jeroen van der Leeuw, Jan Scholtes.*

NICI, University of Nijmegen. *Gerard Kempen, Koenraad De Smedt, Theo Vosse.* (From 1992, Leiden University).

AI Laboratory, University of Brussels. *Luc Steels, Walter Van De Velde, Bernard Manderick, Jo Decuyper, Piet Spiessens.*

Linguistics Department, University of Antwerp. *Steven Gillis, Jef Verschueren, Jan Nuyts, Georges De Schutter.*

Politechnico di Milano. *Marco Somalvico, Vincenzo Caglioti, Vittorio Maniezzo, Domenico Sorrenti, Lorella Colombini, Graziella Tonfoni.*

Apart from these groups the following individual researchers are associated with the proposal as subcontractors: *Chris Turk* (CT Consultants), *Robin Clark* (University of Geneva), *Gerard Wolff* (University of Wales), *Hermann Ney* (Philips Aachen), *John Nicolis* (Patras), *Claudio Rullent* (CSELT, Torino), *Cesare Oitana* (Gruppo Dima, Torino), *Andreas Dengel* (ALV project, Kaiserslautern).

2 Aims of the Project

2.1 New learning techniques for natural language processing

The main goal of the SHOE project is to force a breakthrough in the development of Machine Learning techniques for the extraction of hierarchically structured linguistic knowledge in a selection of crucial linguistic domains. The specific objectives are to characterize learning paradigms in terms of the amount of information they need, to investigate how far unsupervised techniques can actually get, to develop hybrid

supervised-unsupervised techniques which derive the additional information they require automatically from cross-hierarchical correspondences, and to characterize the situations in which such hybrid approaches can effectively operate in an unsupervised mode.

State of the art Speech and Language Processing Systems make use of an array of different linguistic knowledge bases: phoneme and diphone databases, phonetic and phonological (including phonotactic) rules, a lexical knowledge base containing heavily structured syntactic, semantic and pragmatic knowledge, word and sentence structure grammars, domain and user models. In existing language processing systems the acquisition of these knowledge bases is largely by hand (a technique which has been jokingly called "learning by brain surgery"). The problems with this approach, which is dictated by necessity, are evident: linguistic knowledge acquisition is a serious bottleneck for the development of language processing systems (errors, inconsistencies, long development cycle), and even worse, the work has to be redone (often from scratch) for every new language, application domain or application area. We will refer to these two bottlenecks as the *knowledge acquisition* problem and the *knowledge adaptation* problem. Existing Esprit Research and Development projects like SUNDIAL and PLUS are good examples of the absence of automatic acquisition of linguistic knowledge bases.

It is our conviction that basic research in the use of machine learning techniques for the extraction of linguistic knowledge will alleviate these problems for future research and development projects on Speech and Natural Language and Information retrieval by providing tools and techniques for (semi-)automatic extraction and adaptation of linguistic knowledge bases. The main advantage of this approach is the fact that the same techniques can be reused for different languages and different sets of primary linguistic data. On the other hand, learning techniques which require too high a level of human supervision will not realize this potential. Hence we are primarily concerned with investigation of unsupervised learning.

Constructing and interpreting hierarchical structure is a basic activity in all areas of linguistic processing (interpretation, generation, evaluation). Abstraction (generalization) hierarchies of rules and representations are a salient feature of phonological, morphological, and lexical (syntactic, semantic and pragmatic) knowledge bases. Not all existing learning techniques are equally well suited to extraction of hierarchical knowledge from primary linguistic data (speech and text). Our method of achieving the main goal of this project is therefore to evaluate existing symbolic and "subsymbolic" approaches to Machine Learning on their merits in extracting linguistic knowledge in a selection of linguistic domains. The analysis of the results of this comparison will result in a deeper theoretical insight into which properties of existing techniques are necessary for the task at hand. Another probable result will be the development of new, possibly hybrid, learning techniques for the extraction of linguistic knowledge.

Then there are a battery of relevant techniques from outside the conventional Machine Learning fold where their wider applicability has never been considered: analytic techniques (splines, regressions, adaptive prediction), classification techniques (clustering, clumping, seriation), and others.

In some of these cases there are also theoretical characterizations of the efficiency of the techniques and the assumptions necessary for them to be effective. In some of these cases the wheels have been reinvented without making the connection to the existing results and theory.

Recent work in Machine Learning of Natural Language has rediscovered that language is so highly structured that even simple clustering techniques, like those used

in Quantitative Linguistics and Automatic Classification, can be used to discover linguistic classes in relation to their contexts. This is a form of unsupervised learning, and in the Neural Network context it arises as self-organization. As unsupervised, or automatically or implicitly supervised, learning is clearly the paradigm which would normally be preferred in practical ML applications, the project aims to characterize the relationships between different ML techniques, different (sub)domains and different levels of domain complexity.

The domains of natural language and formal languages have had a considerable impact on the development of complexity theory as well, and this along with psycholinguistic results has fed back to shape Linguistics in massive new ways. In particular, the Chomsky hierarchy is used to characterize the complexity of languages in terms of formal machines. Natural language is arguably at least a context-free language according to this method of characterization. Psycholinguistic evidence suggests that children don't get or use much negative feedback on their language production, whilst formal results in language learning show that context-free languages cannot be learnt without implicit or explicit negative information.

Unsupervised learning techniques exclude both guidance by a teacher (implicit negative information) and decisions from a critic (explicit negative information). This is the mode of learning that children apparently use. So we have a paradox.

The solution of Chomskian linguistics is to suppose that there is an innate universal language plus an innate parameter setting mechanism. Computationally, we would prefer to formulate it a little differently, although the parameters that have been discovered provide useful information about the nature of language, and some consideration of machine learning techniques for parameter setting have and are being explored by contractors or subcontractors within the consortium.

A more general solution is to observe that the cognitive restrictions imposed by the human learner have shaped natural language so that it is not just learnable by the human learner, but defined by the human learning mechanisms. This has three implications which are fundamental to the SHOE project. First, the restrictions on natural language are not necessarily those which are produced by the traditional restrictions on formal machines. Second, restrictions in human sensory-motor processing will be similar across modalities, and common cognitive mechanisms suggest that a bottom-up unsupervised self-organizing approach should be able to recover structure in the levels which correspond to pre-linguistic cognition. Third, the language level is characterized by more than mere syntactic structure, and the essence of language is in fact the interrelationship between structures, so that semantics is defined and grounded through interhierarchical learning processes.

These observations lead directly to the most distinctive goals of SHOE: We intend to characterize the complexity of subproblems within language learning in terms of the learning paradigm and the restriction parameters; we wish to apply the same techniques in several different subdomains of language; we aim to explore just how far up the individual hierarchies self-organizing techniques are viable; we wish to examine to what extent unsupervised techniques are able to discover interhierarchical relationships; and we expect to be able to derive sufficient information from these interhierarchical relationships to be able to provide criticism which will allow us to employ supervised techniques to learn at higher levels of the individual hierarchies. This should ultimately lead to more powerful hybrid unsupervised techniques for multiple interrelated hierarchies, which are able to make explicit and employ the implicit negative information embodied in the semantic and other interhierarchical relationships.

The question of feedback from processing at a higher level to influence or even modify the original classification at a lower level is also of considerable importance. Apart from the unsupervised/supervised distinction of connectionist techniques which separates out self-organizing networks, this is the other main distinction between the techniques: what is the nature of the recurrence between levels. And in non-recurrent networks, the question becomes one of how to control the learning parameters in the different layers so that more complex results can be learnt at higher levels without interfering with the successful processing already learnt at lower levels. The corresponding questions for other approaches remain largely unexplored.

Our answer is that each level should operate largely by self-organizing means, but that contextual input from higher levels, and at higher levels from parallel hierarchies, provides the additional input required to determine the correct output of a level.

Our ultimate aim is clearly very ambitious, and whilst there is good evidence for the hypotheses on which our goals are based, the approach is not without risk. For this reason, we wish to take a more exploratory path in undertaking this project, and aim to establish a good understanding of what is going on at one level before proceeding to something more concrete. What makes our goals seem achievable is that we have chosen a restricted domain and aim to explore the lower levels of this domain thoroughly before specifying the follow-on more precisely.

2.2 A matrix of language and learning technologies

We propose a three dimensional lattice of research, with several ML techniques tried on a few NL subdomains, in a number of different languages. We plan initially to undertake narrow probes parallel to each of the axes of this space, which will then guide us in determining which subspaces are worth exploring further. This will allow us not only to establish the extent of the viability of different machine learning techniques, but provide us with our new metric for the characterization of the complexity of languages and problems.

Such an approach addresses the problems of efficiency and overkill which are common in the application of Connectionist and Machine Learning techniques indiscriminately to different problems. Our preliminary analyses leads us to believe that much work in this area uses more powerful, and thus less efficient, techniques than required for the target problem.

We also take into consideration that Machine Learning techniques may be classified on a number of spectra, including particularly symbolic vs connectionist, rule-based vs case-based, and unsupervised vs supervised. Within this space we may characterize the techniques further in terms of the nature and source of the examples and the criticism, the representation of the rules or cases, and the nature of the interactions permitted between the various levels within the hierarchy developed by the technique: consider the relative focus on interactions within and between the layers of a neural network in the self-organizing and back-propogation paradigms, the different ways of representing the same information in decision-trees and rules, the contrasting use of examples in the case-based, genetic and explanation-based approaches, the imbalance between the demands on the user and application in the clustering and the concept learning paradigms.

Not every technique is applicable in every domain, but there is considerable scope for exploring the mating of learning approach and application level. The main restriction is that we don't plan to proceed above the phrase level within the timeframe of

the present application. However, in view of the number of learning techniques and low-level language domains, our approach to the matrix will be one of selecting key algorithms and subdomains, and exploring the other dimension with appropriate examples representative of the different classes of procedure and problem. These probes will be used to heuristically guide further work on the matrix, and are conceived as providing support for the more ambitious probes into the higher dimensions of multi-hierarchical language learning which will be specified for possible follow-on projects.

2.3 Interdisciplinary analysis of existing techniques

Machine Learning has been studied in relation to Natural Language since the earliest days of Artificial Intelligence. Some of the leading names in Machine Learning have been associated with this area, including both the founding editor of the Machine Learning Journal and his successor (Pat Langley and Jamie Carbonell resp. - see their joint paper in MacWhinney, 1987). For a review by one of the proposers, see the Preface to the 1991 AAAI Spring Symposium on Machine Learning of Natural Language and Ontology. For a different perspective see Jane Hill's entry in the Encyclopaedia of Artificial Intelligence.

Machine Learning of Natural Language is, however, intrinsically interdisciplinary, and relevant work also appears in guises such as Quantitative Linguistics (see the QUALICO Proceedings edited by one of the proposers) and Automatic Classification and Thesaurus work (see the entry by Karen Sparck Jones under the latter heading in the Encyclopaedia of Artificial Intelligence). Then of course there are Psycholinguistics and the other obvious areas of Cognitive Science. In fact, it not infrequently arises that a 'new' Machine Learning technique is a reinvention, with a new terminology, of one that has been studied elsewhere.

Another primary objective in proposing SHOE is therefore to bring together researchers from the various disciplinary backgrounds impinging on Language and Learning, seeking to save Machine Learning the reinvention of any more wheels through a comparative examination of techniques from all of the relevant areas, in the context of problems relating to language and speech.

Here we are bringing together expertise not only in Machine Learning and Natural Language from an Artificial Intelligence background, but an interdisciplinary team with backgrounds in Language and Learning from across the Cognitive Science spectrum, but specific expertise in the area of learning in relation to language. The manpower in the partnership is fairly equally divided between backgrounds in Artificial Intelligence and in Cognitive Science, and similarly between primary experience in the area of Learning and in the area of Language.

A subsidiary goal in this respect is the training of researchers in this field. Machine learning of natural language is a scientific discipline that is seriously underdeveloped in Europe. Research is mostly done by isolated individuals in computer science, psychology or linguistics departments. The present consortium will be the first interdisciplinary group in Europe working in this area. It is expected that collaboration in a basic research action will provide a platform for interdisciplinary cooperation, and joint critical mass expertise will become sufficient to compete at a world level, and to interact more effectively with the groups that are emerging in the USA. For this reason, special attention will be given to training of the partners in order to reach a common understanding of the issues and the state of the art.

2.4 Technical and commercial spin-off

Machine Learning of Natural Language is clearly a long way from commercial viability, if one thinks in terms of the HAL of 2001, who indeed acquired his language capability through learning from a teacher. 2001 is a good target date for such a HAL, but 1991 already presents opportunities for the commercial application of Machine Learning in various specific areas within Natural Language.

Whilst much Linguistic and Artificial Intelligence research in Natural Language has concentrated on the higher level subtleties, the SHOE project has deliberately restricted itself to the lower level practicalities. This approach not only allows laying firm foundations for higher level work, from which future forays can proceed on a far more firm and informed basis, but it offers the possibility of developing applications with the promise of immediate commercializability.

One aim of the project is to characterize more precisely the nature and complexity of both learning algorithms and subproblems within the language domain, but this also will lead to the discernment of the optimal approach to the application areas where this technology is clearly already capable of providing practical advantage. These areas include particularly OCR (optical character recognition), speech, and machine translation, and indeed the technologies applicable are similar, and appropriate for similar reasons.

At the lowest level, character recognition with a bitmap template is clearly primitive and unintuitive. It is incapable of recognizing different fonts, and identifying the font. It is incapable of recognizing an unknown font from the relationship of loops and lines. Just as the bitmap has outgrown its useful life as a printing technique, character descriptions, and learning of high level font descriptions promises to avoid the problems where variations in the bit image which are scarcely discernable to the human eye (and cortex) result in unacceptable multiple percent error rates in even single font text.

This approach is analogous to the SR (speech recognition) use of models of speech production to provide appropriate higher level descriptions of the phones to be distinguished and the filtering appropriate to the task. It is also appropriate to the new generation, pen-based, note-pad computers.

At a slightly higher level, we can observe that in fact we do not clearly identify every spoken sound with an unambigous phoneme. Nor do we necessarily identify every handwritten character individually. Rather it is well known that in both hearing and reading, we have a gestaltist recognition of words. Overall shapes, the context of the beginning and end of the sentence, and local consistency provide cues. Even an unknown name obeys rules at the subword levels (a character ambigous between say 'c' and 'e' could be decided on this basis depending on whether a vowel or a consant would be more likely). The methods of QL (Quantitative Linguistics) can be viewed as a form of learning which provides precisely this sort of information.

At a level just a little higher again, we may still not be completely sure of a word, perhaps one of those unstressed little words like 'if' or 'of'. It is possible to create a lattice of the possible readings, perhaps with information about the lower level features and probabilities, and to seek out a consistent parse through this fuzzy matrix of possibilities.

Note that the problems of homonyms and (near) synonyms, and the imprecise correspondences between concepts of different languages, clearly make the use of fuzzy and lattice techniques appropriate for MT (Machine Translation) as well. Again, we don't need to wait till we are in a position to make a complete parse, but the sort of

local information which can be developed using QL techniques is also applicable.

A second area where commercial offshoots of SHOE may be expected concern our development of a library of standards, algorithms, test data and subdomain application solutions. Whilst it would be expected that the library would be licenced to research groups for a nominal charge in the early stages, as it develops it will clearly become increasingly valuable and might be commercialized at some point. More than that, it will be of immense benefit in developing commercial language learning applications.

The aims for this library are of course a mix of technical, commercial and scientific aims:

Standardization: for reproducibility (too many learning algorithms are tested inadequately on undisclosed and often confidential or proprietary data); for mixing and matching (so that new techniques can be tried with old datasets and vice-versa); for presentation (so that new algorithms can be compared with old through use of benchmark problems); for terminology (terms like supervised and unsupervised, critic and teacher, applied to algorithms, and tagged and untagged, positive and negative, applied to data and examples, define a space in which techniques can be accurately placed and thus characterized); for modules (so that standard algorithms can be embedded in multiple applications).

Dissemination: so that the tools and datasets can be widely used, and the above standards established; through adaption and reimplementation of existing algorithms and techniques from many disciplinary sources; through the collection of datasets and simulations which are portable; through placement of materials into a shareware environment which provides more protection (and rewards upon commercialization) and less restrictions (including avoiding the requirement of having special data collection hardware such as speech or video boards).

Analysis: so that the characterization of the power of different techniques, and the interrelation of the different dimensions of the space of possible algorithms, can be made in a way which gives us an index of the complexity of the problems and the nature of the advantage one algorithm or feature of an algorithm has over another; so that the complexity of language and language subdomains may be characterized in a more practically significant way than through abstract machines and automata theory.

2.5 Summary of aims

- To explore the range of efficacy of unsupervised hierarchical learning.
- To characterize the complexity of linguistic subproblems using learning.
- To develop standards to allow the cross-application of technique & data.
- To provide a simple but adequate common intermediate representation.
- To examine techniques developed in other disciplines in this domain.
- To compare competing learning techniques and their applicability.
- To learn from inter-hierarchical data and discover interrelationships.
- To apply learnt inter-hierarchical relationships as supervisory input.
- To use intra-hierachical feedback to improve lower level classification.
- To extend the unsupervised paradigm for multi-hierarchical tasks.
- To transfer technology through dissemination of data and algorithms.

3 Background of the Project

The general theme of the project is an investigation into the application of machine learning techniques to the acquisition of linguistic knowledge bases from linguistic corpora. This is necessary if we hope to solve the knowledge acquisition and knowledge adaptation bottlenecks in current speech and natural language processing systems using automatic acquisition techniques. As the problems and algorithms take many different forms, a corollary is that we will also have to address issues of representation, operationality and standardization. This section describes the current research efforts of the different partners, which will act as a scientific background to the project proposed.

3.1 Learning methods and complexity

The specific theoretical basis of the project results from a critical analysis of the way machine learning techniques are employed and the evidence and criticisms from the cognitive sciences from the point of view of human learning and cognition. In particular, it is apparent that the traditional complexity hierarchy of languages, based on formal machines, does not adequately reflect the nature of the human cogntive restrictions which determine the structure of natural languages. Furthermore, the application of black-box machine learning methods (often connectionist) may result in use of a paradigm which is unnecessarily complicated and better suited to more complex tasks.

Learning algorithms can be divided up in several ways, but we wish here to consider them primarily in terms of the paradigms under which they may be employed - whether they require positive and negative examples, can generate and test their own examples, or can work with naturally occurring input. We characterize them in terms of the use of a teacher, by which we mean provision of examples in a helpful order, and/or a critic, by which we refer to availability of positive/negative judgements or multivalued classifications. We also note that in complex environments the effect of teacher and/or critic may sometimes be derived naturally within the system.

The dozens of learning algorithms may also be characterized according to the area of science in which they, or their underlying metaphor, originated. Related to this we may also consider the intrinsic parallelism, or potential therefor, in the method. Grouped according to the nature of their use of teacher and critic, we can then proceed to compare effectiveness on the same data sets.

Conversely, exploring different problems and the paradigms in which they are solvable will provide a new form of complexity hierarchy. In particular, we focus in this project on the extent to which problems arising in a broad natural language connection can be achieved using an unsupervised paradigm, without teacher or critic. We proceed on the basis of a hypothesis that the lower levels of language have sufficient intrinsic structure that such methods are capable of clustering out the various categorizations we require. The phrase level is where we expect to find the limit of this methodology in the syntactic domain, the entity level in the ontological.

Powers (1984, 1989) has already shown that word classes and associated rules may be learned either by statistical clustering techniques or self-organizing neural models, using untagged data, thus achieving completely unsupervised learning.

Powers (1991b) applied his earlier work down below the word level, and proposed that such techniques were viable only up to the phrase level, whilst above the phrase level the intermodal interactions (semantic correlations between the ontological representation and the linguistic representation) provide additional cues which are necessary

to continued unsupervised learning, and allow using supervised learning techniques with implicit teacher and critic.

Note that Turk (1988) has proposed a model of anticipated correction, which is in accord with the Powers (1984) hypothesis of a logical separation between recognition and production allowing the recognition grammar to act as critic for the production grammar. In the Powers and Turk (1989) implementation of this model, thresholds were used to control movement of grammatical rules from the recognition grammar into the production grammar, with a piagetian assimilation/consolidation process compiling cases into rules. Viz. the recognition grammar can be regarded as more case-based and the production grammar more rule-based. In fact, in some methodologies, the progression can be automatic and virtually seamless.

Marcus (1990) and Powers (1991b) have independently shown that forgetting the idea of top-down 'constituent' parsing has some advantages, and Marcus and Magerman (1991), working with a tagged corpus, have shown that a 'distituent' grammar can be developed, at least up to the phrase level.

Powers (1984, 1989) has implemented all his models with the modality as a parameter, so that the same techniques can be applied in the ontology, phonology, morphology, and some other subhierarchy. To allow this to be explored he has designed a robot world model (with Hume, 1984) and shown (with Chan, 1988) how simple noun and verb semantics can be learnt. Reapplication of the approach to the orthography to word level, Powers (1991b) showed that phonological features and categories are so strongly intrinsic in language, that they may be clustered out from the orthography (by analysing a full dictionary containing only ascii word representations).

3.2 Interdisciplinary insights and cross-application

A primary aim of this project is to critically examine models and approaches used for automatic discovery of structure in other disciplines. One of the main areas where such techniques have been used is in the study of literary styles and meaning ranges. Further afield are some of the metaphors from physics and biology. The latter are being studied particularly in Brussels, Tilburg and their associated partners, and the former in Trier.

Rieger (1977, 1979, 1982, 1983) has developed fuzzy techniques for clustering with a view to characterizing word semantics. Recently this has lead to interdisciplinary work (Rieger 1989, 1990) on knowledge acquisition from corpus data.

Daelemans (1991) describes an approach in which a hierarchy of phonological rules is learned by using a selectionist learning technique. A corpus of linguistic data constitutes the environment to a population of linguistic rules that recombine according to fitness (operationalised as the proportion of the 'environment' the rule correctly describes), and using the idealised genetic operators crossover and mutation. In current research, a variant of this approach, genetic programming (Koza, 1991) is applied to the same and other phonological problems. The advantage of the latter approach is that the selectionist algorithm works on symbolic structures that can be inspected, interpreted and reused.

Wolff (1991) shows that unification and compression techniques could not only be used as a language learning mechanism, but show possibility for more general use in cognitive and computational modeling (see also Powers(1989,1992a,b)).

Recent proposals from Nicolis (1991) on applying chaos theory to linguistic processes are supported by remarkable explanations of neurological phenomena, in partic-

ular that the thalamocortical pacemaker acts as a multifractal strange attractor, with the result that natural restrictions are imposed on the grammatically legitimate words, the significant key features of a pattern, and the set of "moves" available. This allows one to limit attention to a few key features which code disproportionately high levels of information.

As the effect of this is similar to that of the statistical processes used by Powers (1991) in syntax, and the quantitative linguistic techniques used by Rieger (1990), a comparison of the techniques and evaluation of the wider potential of the chaos theoretic model will be undertaken. Also, relationships with biologically motivated learning algorithms will be explored with other partners.

3.3 Representation, operationality and standardization

The choice of representation has an impact on the operationalization of an algorithm and interchange of modules.

The traditional AI languages, LISP and PROLOG, provide general structures and mechanisms which are useful in this respect, and it is envisaged that logical terms will be a primary data structure, both for processing and interchange between processes. The unification procedure has already been demonstrated by Powers (1984, 1989) to be useful in language learning in the differential minimization procedure for unifying subtrees. This technique is closely related to the techniques of feature unification which have become increasingly standard in Computational Linguistics, and are being actively developed and pursued in different ways by many of the Partners and associated groups and projects (Uszkoreit 1989, Siekmann 1987, Rollinger, 1991, Smolka, 1988).

Daelemans (1987a,b; 1988, 1989, 1990) has shown that approaches from object-oriented programming like encapsulation, polymorphism and inheritance are also appropriate knowledge representation primitives to represent the pervasiveness of hierarchical organization in different areas of linguistic knowledge. Phenomena like blocking of regular rules in the presence of exceptional rules, degrees of markedness and productivity of rules and representations, and regularity - subregularity - exception dimensions follow automatically from these representational primitives. See the special issue of Computational Linguistics (Daelemans and Gazdar, eds. 1992) for an introduction to and a sample of research about these issues. This implies that learning techniques, to be useful, should be capable of deriving such hierarchically organized structures or rule sets from primary linguistic data.

3.4 Ontology and grounding

One of the complaints raised against the Artificial Intelligence approach to Natural Language is that we do not have a viable concept of understanding. Searle's Chinese Room argument and Harnad's Symbol Grounding argument are supposed to show that traditional Turing test definitions of understanding, and conceptual structure approaches to semantics, are inadequate to warrant the use of terms like understanding and thinking in any more than a weakly metaphorical sense.

Whilst much of their argument is NOT well taken, and in Computer Science we don't necessarily want to make great philosophical or psychological claims to validity, this much is clear: Work in which so-called natural language is translated into a so-called representation language comes no nearer to understanding in any sense, but leaves the work of understanding to some as yet unspecified homunculus. Searle takes

it for granted that the problem is that the homunculus must be responsible for the understanding or "mind" of the system. Harnad takes it for granted that there are aspects of the biological perceptual and cognitive apparatus which are inseparable from real understanding. Turing takes it for granted that realistic communication is possible in a system which has no real grounding in the world.

We adopt a middle ground. We are happy to work in a simulated or limited application environment in which semantics can be operationally defined. We want to use a representation language not as an end of the understanding processes but as a basis for linking the linguistic and ontological modalities. And where possible we want to use the same mechanisms and representation formalisms in all levels and hierarchies within the framework of all modalities. Powers (1983, 1984, 1989) has argued this extensively, and has provided the robot world framework in which to explore the simultaneous learning of ontology and syntax, as well as the learning of the semantic interrelationships.

As discussed in the following section, the physical metaphor is so pervasive in language that it is arguable that a naive physics background is necessary in order to be able to correctly use even the most basic words, such as prepositions, in natural speech, in even a specific application domain in which a controlled system provides the ontology. Language is productive, and such a system cannot hope to predict all usages within the system without having some grounding in the system which underlies human language usage.

The logical extension of the symbol grounding problem is Harnad's robot who is able to pass the Total Turing Test and be totally indistinguishable from a human. We will however content ourselves largely with simulations taken to a level where it will be reasonable to expect that some of our concept learning work will prove useful eventually in computer vision, or more accurately, integrated AI systems.

Rollinger (1991) in LILOG has been concerned with the study of how ontologies can be developed and linked with the concept based representation languages, whilst Powers (1989) has been working with a parsing model using tree-like representations of objects and actuators, views and actions. Emde (1989, 1991), both in his doctoral research and as a project leader in LILOG-OS has been concerned with non-conservative knowledge revision and concept learning.

3.5 Sensory-Motor

An important aspect of the proposal is that work on grounding should include the automatic learning of an ontology from a simulated robot world. Although it is not envisaged that we will seek to make use of an actual robot, Milan has expertise in robotics, multirobot architectures, and control of robots using natural language, and has a mobile robot available.

The contractors will also not directly seek to work with Speech and Vision at the perceptual level, except insofar as subcontractors and independent projects within the partner institutions are interested in applying techniques to applications at this level, and contractors' experience in specific learning techniques, phonology or concept learning is relevant.

A number of commercial concerns involved with speech and machine translation are working with techniques in or akin to machine learning, and wish to have some involvement in the project. In addition a separate group within the coordinating contractor is working in optical character recognition. No vision group has sought links with the

project and it is not envisaged that problems of visual perception will be dealt with other than at the level of the simulation, Philips (Aachen) and CSELT (Torino) are working in Speech Recognition, DFKI (Kaiserslautern) in Optical Character Recognition and Gruppo Dima (Torino) in Machine Translation.

It is proposed that the coordinating contractor and the project manager will, through exchanges of personnel, data and techology, pursue a small cluster of related tasks linked with these areas of immediate commercial potential.

Dengel et al. (1988, 1990) are working on using higher level knowledge sources (including large lexicons, presentation, layout, knowledge about logical document structure and domains) to assist in the reading and understanding of business letters. In addition they are studying word-based recognition and geometric knowledge acquisition for automatic logical labelling of document blocks. They will be primary participants in the parts of the commercially oriented workpackage concerned with character recognition and reading, and will seek to explore interchange of learning and analysis techniques with the commercial speech labs - who will be subcontracted to the DFKI.

In a sense this area, exploring speech, reading and translation, is the least essential to the project, but the partners feel it is important also to provide an avenue for the severe testing of learning techniques offered through strategic research involving commercial subcontractors (who will also be contributing to the general progress of the project through participation in workshops, training etc.) It should be noted that these subcontractors would have been interested in Associate Partner status and responsibilities, but practical considerations such as distance, size and major involvement with other national and international projects has precluded this level of involvement. It is, however, expected that their demands on the resources of the project will be more than balanced by their contributions.

4 Description of the Project

In this section, we will describe the project we propose in the context of traditional issues in (automatic) natural language acquisition research.

4.1 Bootstrapping and negative information

One of the fundamental problems in the machine learning of natural language (as well as in theories about natural language acquisition) has been christened the bootstrapping problem. Cast in psycholinguistic terms: how does the child break into the linguistic system? What knowledge does he bring to bear on the task of starting to construct a formal linguistic system? The problem can be reexpressed in terms of resolving the discrepancy between the input a language learner receives and the result of the learning process (viz. a rule system - irrespective of what form it takes).

The difficulty is that there is no direct relation between the information contained in the input and the information that is essential for the output of the learning process: there is no grammatical marking in the input or in the context in which it occurs. Pinker (1984, 1987) argues that even if we take into account perceptible input - information such as the set of words contained in a sentence, their order, prosodic properties (intonation, stress and timing), phonological properties of words, the semantics of words as inferred from the nonlinguistic context and pragmatic information inferred from discourse context - the problem remains as to how a child arrives at the

rules of the adult language which is couched in terms of syntactic categories, grammatical relations, cases and phrase structure configurations not explicitly indicated in the input.

Various solutions to this problem have been proposed:

Correlational bootstrapping basically proposes a form of learning on the basis of observing distributional constancies in the input. Formal syntactic categories are constructed by analyzing the copredictive structure of structural properties per se, without an appeal to semantic properties to bind them together into categories (Maratsos & Chalkley 1981, Maratsos 1982, 1983). Slobin (1985), Pinker (1987, 1989) as well as Maratsos (1990) argue against such a bootstrapping operation precisely on the basis of a lack of semantic integration.

Prosodic bootstrapping holds that the acoustic packaging of the input language is such that it provides markers of the major syntactic units. These regularities can be used to infer the syntactic structure of a sentence (Morgan & Newport 1981, Wanner & Gleitman 1982, Hirsh-Pasek et al. 1987, Kemler Nelson et al. 1989).

Syntactic bootstrapping holds that the syntactic structure per se constitutes the entry into language structure. A small amount of distributional learning is sufficient to yield correct grammatical rules because strong innate restrictions on possible grammars severely constrain the induced syntactic forms (Grimshaw 1981, Pinker 1982).

Semantic bootstrapping basically claims that semantic notions are used as evidence for the presence of grammatical entities, esp. syntactic categories, grammatical functions, cases, grammatical features and tree configurations (Wexler & Cullicover 1981, Macnamara 1982, Pinker 1984, 1985). It presupposes four basic assumptions (Pinker 1987): (i) meanings of content words can be derived independently from the context; (ii) the semantic representation of a sentence can be constructed on a contextual basis; (iii) the semantic inductive basis, the formal grammatical categories as well as the mapping rules between them are innately given; and (iv) the semantic-syntactic correlations hold in the 'basic sentences' of the language which (exclusively) form the initial input to the language learner; in other words, the semantic elements are sufficient conditions for use of the syntactic symbols in basic sentences.

It appears that the various forms of the bootstrapping operation depend on the presence of prewired (innate) knowledge and/or procedures. This brings up another basic problem in natural language acquisition, viz. the notorious controversy known in the psycholinguistic literature as the nature-nurture debate. One can safely assume that children construct an internalized grammar by using incoming language data together with innate (linguistic) knowledge to formulate hypotheses about possible grammatical rules. But what are the relative contributions of innate (or a priori) knowledge and the structure of the input language to the learning process?

Taking Pinker's semantic bootstrapping hypothesis (see above) as an example, there is a fair amount of innate knowledge involved even with respect to the semantic inductive basis of the formal syntactic categories. There is a growing body of evidence (see e.g. Slobin 1985) showing that the initial semantic categories of children learning different languages show many intriguing similarities. These similarities can be accounted for if it is assumed that the way children conceptualize and classify the elements of their experience is not free to vary arbitrarily, but rather is shaped and constrained by inherent properties of the human perceptual and cognitive system. But in a brilliant paper, Bowerman (1989) shows that children – even in the earliest stages of language acquisition – are highly sensitive to the way the words, grammatical forms and construction patterns of their language are used. Thus, although learners may sometimes

match language forms to concepts generated independently of linguistic experience, they are also able from the earliest stages of language acquisition onwards, of building language-specific categories by observing the distribution of forms in adult language and making inferences about the categorization principles that inderlie them.

Any system that aims at natural language acquisition will ultimately have to deal with the problem of prewired versus learned information, a problem that turns up in all strata of the linguistic system. For instance, following Chomsky's influential arguments for an inborn 'Language Acquisition Device' (Chomsky 1965) the controversy has been focused on whether there is innate knowledge of syntactic structure and on what form it takes. A similar problem turns up with respect to learning phonology: Dresher & Kaye (1990) adopt Chomsky's (1981) principles and parameters model for learning the English stress system. They assume a learning process that consists of fixing eleven parameters which have been shown to underlie stress systems and which should lead the learner to the postulation of the system from which the primary linguistic data are drawn. The parameters are drawn from universal grammar and thus they are innate (but see Bates et al. 1988 and Hardy-Brown 1983 for a demystification of the latter implication).

One of the main problems for the machine learning of natural language is identified in Langley and Carbonell (1987) where they give a to-the-point critique noting that all existing modeling systems cheat by 'hand-crafting' the input to the model. The input contains exactly the right features for the language being learned, and often only those features, which reduces the learner's search problem. The 'hand- crafting' of the input is analogous to the 'prewiring' of knowledge: the former is functionally equivalent to the latter. This observation is also implicit in Gold (1967) where the result on the unlearnability of context-free languages admits one fundamental exception: 'anomalous input', by which is meant input in precisely the right order to allow some 'subset principle' (Berwick, 1983) of 'parsimony' (Powers, 1983, 1989) or 'simplicity' (Brown, 1968) to operate and specify a unique grammar.

Closely connected to the problem of prewired versus acquired knowledge is the role of the linguistic environment. Learnability theory or formal learning theory shows how assumptions about the language (or class of languages) to be learned, the environment of the language learner (i.e., the information that the learner has to use when acquiring the language) and the learning strategy (or grammar forming mechanism) impose constraints on the learnability of the language. Like all induction problems, language acquisition is difficult because an infinite number of hypotheses is consistent with the finite input sample. However (some of) these hypotheses differ from the correct hypothesis (the target language) in ways that are not detectable given the input sample alone. This problem, also known as Baker's paradox (following Baker 1979) results from on the one hand the need to generalize over the input received (productivity), and on the other hand, the need to block the application of productive rules in particular environments not present as such in the input.

A key factor is the role of negative evidence, i.e., information about the grammatical status of input sentences, or in other words, information about which strings of words are not legitimate sentences of the language to be acquired. If the learner is not provided with negative evidence in any consistent way, as is the case in language acquisition in a naturalistic setting, the problem turns up how to avoid formulating overly general rules solely on the basis of positive evidence (Bowerman 1983, 1987, Pinker 1984, 1989a, b). Whatever solution is proposed for the 'no negative evidence problem' (subtle negative evidence, conservatism, preemption, etc., see Pinker 1989a

for a review and thorough discussion), it should at least show (i) how the learner is constrained to entertain a restricted set of hypotheses that includes the correct one but excludes many others, and (ii) how the learner goes about comparing the predictions of an hypothesis with the input data so that incorrect hypotheses can be rejected.

4.2 Incremental Learning

A second area of fundamental importance is the question of how the data is presented. In the above discussion of bootstrapping, we saw that a fortuituous order of presentation - such as might be provided by a teacher of an oracle - is equivalent to critical input, or negative information, in some sense. But if all the input necessary for specifying the grammar is presented in one go, and we then aim to learn the best grammar to fit that input (where 'best' may include some idea of 'parsimony' or 'simplicity'), we have a similar advantage.

With this 'all at once', or equivalently 'full memory' learning, we have several advantages. It is possible to find an optimal explanation for the given data. If we know we have enough information to decide our grammar, this is itself a piece of critical information. If we have full memory of all input we can make this assumption after each piece of input, and continually totally revise our model at each step to provide an optimal explanation of our input.

An incremental algorithm (Winston, 1983) operates with limited memory and modifies its grammar on the basis of the current input, the currently hypothesized explanation, and a limited memory window into previous input.

We assume in this project that the incremental model is more appropriate to the scale of the language learning domain, and fits better our model of the human language learner. Moreoever, we assume that it is not necessarily possible or appropriate for full use to be made of each single presentation, and that no single input should be given 100% credibility as being grammatical. We therefore also assume that sentence types recur in the input, and the grammar that is learnt is based on the constructs which recur frequently. If the memory window is large enough to contain all significant cases this is equivalent to the full memory model.

We, however, hypothesize that the window is much smaller than the number of cases necessary, and therefore expect that the pure case-based approach will not prove useful in the general case for language learning. What 'small' means in this context can be expressed in terms of the "Magic Number Seven Plus or Minus Two" of Miller (1967). But we must then decide where this criterion bites.

In this project we are looking at learning hierarchical structure, in which the language problems are seen as being decomposable into small numbers of pieces at each of a number of levels, forming a number of logically independent hierarchies (at the lower levels) with interrelationships between them. The number of units retained at a given level, and hence the arity of the decomposition, are bound by a small constant of the order of seven in our models.

The way in which we can most easily see this hierarchical structure being learnt without violating our constraints on memory and bootstrapping is by using clustering or self-organizing techniques whether statistically, symbolically, biologically or physically motivated.

Hierarchical models and theories go under a number of different names in different areas of linguistics, to go no further: e.g. hierarchical, level-based and even non-linear.

4.3 Classification, clustering and taxonomy

Automatic classification research started in the fifties, both with the application of exisiting statistical techniques and the development of new theoretical approaches. Research addressed both hierarchical and non-hierarchical classification, preempted some of the techniques which are today employed under the banner of fuzzy logic, and dealt with problems of non-exclusive and ambiguous membership. Whilst some of the language-related applications in information retrieval have become well known, the usefulness of the statistical approach has been explored in many contexts, from anthropology through cryptography to archaelogy (Sparck Jones, 1990).

Unlike some of the modern reincarnations of classification, in for example self-organization, considerable analysis has been performed on these techniques, allowing a characterization of their soundness in the formal senses, and the consideration of the psychological validity (in terms of strengths and, in particular, weaknesses shared with the approach).

It will be noted that the terminology used is very broad and varies with both application and geographical area. The issue of use of "relevance" information is analogous to the question of "supervision" in Machine Learning. The precise relationship remains to be explored.

When statistical methods are used, it is well known that it is important to ask the right questions and interpret the results soundly. When using statistical methods for learning, the same considerations apply. Work by Powers (1984, 1989, 1991), Rieger(1990) and Nicolis(1990) has provided evidence that the structure which is necessarily inherent within natural language and ontology can be detected by statistical means. Whilst much modern work in Quantitative Linguistics is concerned with studying the semantics and usage of language, and earlier work sought to derive rules of syntax using statistical techniques, the current approach seeks to establish segment boundaries and highly significant class associations.

The Powers(1989) research showed that similar results in learning word classes automatically, with unsupervised techniques from untagged text, could be achieved with radically different techniques: statistical and connectionst. Powers(1991) showed that phonological classifications could be recovered from ASCII words using statistical techniques, and that using four different metrics identical and strongly significant classifications could be achieved.

Once these strong classes have been discovered, they act as natural boundaries for further analysis, and it is natural to define further analysis on a new level using the classifications and boundaries discovered at the lower levels. A particular advantage of the high reliability of these classification techniques, irrespective of their match to linguists traditional classifications, is their stability whilst we are trying to learn higher level clusters.

Traditional approaches in which the levels are imposed or where too much has been attempted at an individual level have proved to be unsatisfactory and unstable. The main contribution of the backpropogation technique in neural networks was the stabilizing effect, but this is originally derived from additional information supplied as classifications of the training data. Incremental techniques allowing more complex examples to be introduced later are also faced by these problems, but few good solutions are available.

But if the teacher needs to supply graded examples, and the critic needs to supply classifications, and we still have problems, there is clearly a mismatch between our

paradigm and our problem. And then noise and data errors add even more problems.

4.4 Noise, errors, complexity and the teacher

One of the inherent advantages of statistical and most biologically or physically motivated learning techniques is that they are inherently insensitive to errors or fuzziness in the data. In order for symbolic approaches to handle noise, it is essential to include probabilistic or possibilitic features. In some classical approaches, such as tree-based, case-based, and explanation-based approaches, the use of metrics based on information, distance or usage is quite usual. On the other hand many of the inductive techniques assume perfect data and crisp concepts.

When input is too complex, many learning procedures cannot make any use of it, and indeed the learning process may be damaged. This phenomena is known from the very earliest work in ML, such as Samuels (1959, 1968) to recent work in MLNL, such as Powers (1984).

Psycholinguistic evidence from both babies and aphasics indicates that input which is too complex is still largely understood, but that some of the finer (semantic) relationships are not recognized. Aphasics have even reported that the components they cannot understand sound like noise. For a learning system too complex input and noisy input have a lot in common. When input is just a little more complex than the current level of competence, the difference between noise and complexity becomes apparent, in that it is possible to learn relationships concerning the additional complexity.

This conforms well to the standard maxim of machine learning: You can only learn what you almost already know. Or putting it another way, input is graded according to complexity. On one extreme, there is input that lies within the child's competence. At the other extreme there is input that lies completely outside his competence. In between these two extremes there are various forms of input that are accessible to a certain degree; i.e., a child can understand particular things in the input but not others.

For example, the child can understand the names of the persons mentioned, but not the roles they play (indicated, inter alii, by order in English and Dutch). In this circumstance he can act as if he understood what is said by applying a 'non-linguistic strategy' or 'comprehension strategies' (what makes most sense in the world in terms of who does what to whom? etc.) Or it may be that the child does not understand the input verbatim but the pragmatics make perfectly clear what is meant.

With many learning algorithms, overly complex input has the effect of disrupting the learning process, and even prejudicing already learnt rules.

The bottom up classification approach has the advantage that given arbitrarily complex input only local relationships at one level are attempted by any individual learning component. Only once the component at the lowest level is sending sensible boundary and cluster information to the next level will the learning process at the next level start to develop useful clusters. But we are not dependent on a teacher to supply simple input first, before proceeding to the complex. On the contrary, the strongest classifications give rise to sharp boundaries which break up the input and direct the focus of attention to the adjacent units. The remaining input is essentially treated as noise at that level.

Marcus (1991) and Magerman (1991) have shown that it is possible to build up a grammar from such boundary information. Because it is based on boundaries, it is called a Distituent Structure Grammar in contrast to the traditional Constituent

Structure Grammar. However, their analysis is based on a tagged lexicon, which thus still provides the classifying function of a critic, even though not the directing function of a teacher.

Of course any simple examples which occur in the input, which are already within the competence of the learner, can at most contribute to reinforcing the classifications and rules they already have. And complex examples, which are totally beyond their competence, function largely as noise. The lower level components on which the complex structure is based will still be handled, but the input as a whole will not be 'understood' because there are too many intermediate levels which have not yet been fully specified. On the other hand the learning of the next level of components may still be able to proceed, remembering that we are assuming a learning system which is based on some sort of statistical clustering and averaging processes.

The most useful input is the input which is just beyond the learner's competence. This is the old ML maxim again. It is also an observable fact in child language acquisition. But the reason seems to be that in some way the whole can be 'understood' with the help of the context. Thus we need to distinguish between input which is strongly structured (so that learning occurs within a single hierarchy of such learning levels) and input whose structure is only discernable with outside help.

Here we propose (Turk (1988,1990), Powers and Turk (1989)) that such outside help is available from implicit sources in typical language learning situations. In particular we note that comprehension always leads production, and suggest that our recognition grammar must play a role in our learning to produce grammatical sentences.

4.5 Implicit and Explicit Teacher and Critic

When a child asks for something and is not understood, he doesn't normally get what he wants. When he tries to do something, but doesn't follow the correct procedure, he doesn't normally get the effect that he wants. If he pulls at a cupboard door without turning the handle, for example.

This is negative information. But on its own it doesn't help much to correct the request or behaviour. What is need is some mechanism which not only gives yes/no correctness information, but which *focuses* attention on the incorrect segment.

A typical child's picture book has sentences like "Tom can skip. Tom can jump. Mary can skip. Mary can jump.", along with the associated pictures. This sets up a paradigm in which differences, specifically unknowns in one hierarchy cooccur with unknowns, new words or situations, in the other. We have already explained how classification techniques can ignore the noise and focus on what is just beyond their competence. There are also focusing effects which arise automatically in the situation. The coincidence of unknowns is one. So is pointing or even just focusing or orienting. This is known as the deictic effect.

The boundary and pointing function within single hierarchy learning can also be regarded as a deictic effect. Huey (1908) showed that readers speed up and slow down at the beginning and end of sentences. Speed reading techniques are based on wholistic treatment of paragraphs, sentences and words. Powers' (1989) classification program detects sentence boundaries first (punctuation), then phrase boundaries (articles and the like), etc. These functional units have the least semantic value, that is they are primarily related to the hierarchy they are found in, and can thus be said to have intrahierarchical roles.

But these units draw attention to (deixis literally means point to) the adjacent

units. The functional units are detected because they are closed classes - thus there are very few words which, like an article, occur in the specifier/determiner position in a noun phrase. However, the semantic content is carried largely by open class words. Nouns, for example, do not have their primary function within the hierarchy, but correspond to interhierarchical relationships: that is structural relationships to and within another hierarchy.

Powers (1983) showed that unknown words could be classified from the grammar for the most part, and multihierarchical context provides additional cues to the grammatical class, as well as to the meaning. Conversely, Powers(1989) suggests that in addition interhierarchical relationships contribute to the identification of appropriate intrahierarchical relationships, and provide additional inputs which may be used by the learning processes in the individual hierarchies.

There is psycholinguistic evidence that children start to learn the structure of their mother tongue very early, even before they are born (Mehler et al., 1988). By the time they get to the stage of their first words, they have already developed some sort of ontology, and are sensitive to the specific linguistic features of the language they are exposed to. They fixate very readily objects offered to them (or carelessly left within reach), and they reach for them and grab them. If the parent uses appropriate words under those circumstances, there will be a strong correlation with the object on which their attention is focused.

There is thus sufficient evidence of focusing to warrant examining the extent to which unsupervised learning can detect interhierarchical relationships. These relationships then constitute additional information which is available to the learning system in the individual hieriarchies, and allow the paradigm to be shifted in the direction of supervised learning without actually demanding explicit supervision.

Focussing has been thoroughly explored by Tonfoni (1990,1991) in her Communicative Positioning Program, which has been successfully used in many classroom and commercial contexts to enhance the effectiveness of writing, reading and learning.

However, an ontological focus does not point solely to an object of focus, but to the whole ontological frame in which the usage occurs.

4.6 Concepts and contexts, paradigm and metaphor

It is a very simplistic semantics which creates a one-to-one relationship between words, even nouns, and objects or actions in the environment. Any word brings with it not only its specific meaning, but a whole framework in which it is to be understood, or indeed a set of possible frameworks which must be unified with the current contextual framework. Virtually all content words must be interpreted relative to the current contextual frame.

Quantitative Linguistics has contributed very considerably to the understanding of the semantic frames associated with words, and Montague Grammar and Situation Semantics seek to provide bases for working with and combining these frames.

Paradigms have long been a successful learning technique, and occur naturally, and we expect usefully, with the focusing effect we have just described. A paradigm exchanges specific concepts in a systematic way, while keeping the same basic context. Conversely, a metaphor depends on retaining the same abstracted concept across a radical shift in context.

Metaphor is not just an esoteric literary form, but underpins the whole fabric of language at every level. This phenomenon has been examined in detail by Clark and

Clark (1973), Lakoff and Johnson (1980), Lakoff (1989) amongst others. Metonomy similarly is fundamental in the way we use language, and is related to our bringing up a whole frame with the use of a single word.

Consider an arbitrary preposition, "in" for example. The fundamental in-ness has to do with space in a physical environment. But we use "in" in time, in relation to thought, and in many other even more abstract contexts. The basic concept remains the same, and is both an influence on and influenced by the way we think.

Metaphor is necessary, because we are seldom (technically never) confronted with exactly the same situation, so in order to apply our experience to new experiences we need to make judgements about whether and where situations are similar and different, and to adapt the most appropriate frame. This sounds very similar to what is going on in case-based learning. The consolidation of many different usages and frames into an abstract concept is the essence of inductive learning.

4.7 Relationships between learning methods

Our approach to learning is one which doesn't require keeping maximum information about every example, unlike traditional work with case-based or inductive learning. Genetic learning can be compared with case-based learning in the sense that we are doing a mix and match amongst the reasonably applicable cases. Backpropogation typically has the effect of learning individual cases while there is sufficient real-estate available, and averaging or consolidating cases which are relatively similar once it is forced to ignore the finer differences.

In the end, our understanding of the world is limited by the resolution of our perceptual system. But in fact we throw most of this information away, and the power of our cognitive system lies in how we structure and relate the information which is ultimately grounded in the sensory-motor. Phonology concentrates for example on when sounds which are actually different are interpreted as being the same, and the unique way in which this is done for every different language and dialect.

In the end there are so many insignificant differences which do make it through our perceptual system, that it becomes totally unreasonable to deal with individual cases. But once we have made the emic classification of the etic data (Pike 1947, 1954, 1977), we can treat these classes as individual cases.

For this reason, we are paying special attention to the biologically and physically motivated paradigms which deal with this reality, whilst continuing to explore the symbolic paradigms to the extent to which we are dealing with distinguished emic classes.

4.8 Complex Dynamics for Natural Language Processing

These new ML paradigms are being developed with the inspiration of concepts and theories from the natural sciences (mainly biology: evolution by natural selection, neural networks, but also from physics and chemistry). These theories involve the massively parallel interaction of a large number of relatively simple elements (e.g. units in a connectionist network, or individuals in a population), and view computation as a complex dynamic system. The new paradigm has been called the *complex dynamics paradigm* (Steels, 1988), and involves metaphors that are markedly different from the

ones common in the symbolic paradigm of AI which were based on concepts and theories from logic, philosophy, linguistics and psychology.

We are especially interested in learning (or rather *self-organisation*) of linguistic processes and representations within this paradigm. See Daelemans (1991) for an overview of existing biological and chemical metaphors in Natural Language Processing (NLP).

We are interested in exploring complex dynamics approaches in a number of linguistic areas, and in comparing the results with alternative symbolic and statistical methods. In other words, we want to investigate how far self-organisation can get us in NLP. For this, we picked three target linguistic areas and three classes of complex dynamics systems. Some of the points in this matrix will be investigated during the first stages of the project. In a second stage, the same target linguistic areas will be investigated using newly developed (and possibly hybrid) self-organisation techniques.

4.8.1 Language areas

- **The self-organization of phonological systems.** How can phonological structure and processes arise from phonetic data? Rather than presupposing innate structure we investigate how structures like phonological skeletons, stress patterns and syllable structure can be extracted from phonetic data by self-organization.
- **Evolution of innate syntactic architectures.** Some success has been achieved using biologically inspired methods for parameter setting in principles and parameters approaches to grammar (Clark, 1991), or using molecular interaction techniques in syntactic processing (Kempen and Vosse, 1989). These approaches presuppose a vast amount of innate syntactic knowledge. What we want to investigate in this context is how selectionist and connectionist self-organisation techniques may explain the existence of this "innate" knowledge. How can self-organisation explain not only parameter setting, but also the existence of syntactic processing architectures in the first place.
- **Lexical acquisition from text.** A number of statistical and symbolic techniques have been developed to extract lexical information from corpora (text, tagged text and machine readable dictionaries). The acquisition of lexical knowledge is indeed an important bottleneck in the development of natural language interfaces, especially in systems that rely on lexicalist grammars, and that must be adapted to many domains. We investigate self-organisation techniques to overcome the shortcomings of existing techniques.

4.8.2 Complex dynamics techniques

- **Self-organizing neural networks.** Backpropagation of errors is the most popular connectionist technique. Unfortunately, this technique presupposes supervision, and does not sustain the automatic generation of categories and structure. We are looking at a class of neural networks (mutually excitatory/inhibitory networks) based on work by von der Malsburg, Kohonen and Grossberg (Powers, P:1989; Scholtes, P:1991). Neural data-oriented parsing (Scha, 1990) is another related technique we will explore.
- **Selectionism.** We are also investigating selectionist learning and self-organization in the context of language learning. Techniques include genetic search and classifier systems (Goldberg, 1989), and genetic programming (Koza, 1991).

- **Molecular Interaction**. Self-organization in the unification space model (Kempen and Vosse, P:1989) and variants will be investigated.

The inherent massively parallel character of these complex dynamics approaches makes necessary a major effort on the implementation of these techniques on fine-grained parallel machines in order to obtain efficiency in our experimental setups. The group in Brussels will coordinate the development of tools and high-level languages for parallel implementation of our algorithms on available data-level parallelism hardware.

4.8.3 Self-organization of phonological systems

From research on natural phonological acquisition, we know that this process consists of two stages. At first, items are incorporated in an unanalysed form into the mental lexicon. At some point in the development, a critical point is reached which makes necessary a reorganization in which crucial use is made of the construction of generalizations (Plunkett and Marchman, 1991). In semantic acquisition a similar sequence of events has been hypothesised (Gillis, 1984; Gillis and De Schutter, 1986). The notion of syllable structure plays a central role in the hierarchical analysis of phonological word forms (Wijnen, 1990). Phonological processes and distinctive features also play a role in the self-organization of the phonological system (see Gillis et al. 1988a,b and Schaerlaekens and Gillis, 1987 for phonological acquisition in Dutch). In this sub-project we will investigate which properties a learning algorithm should have to self-organize into a complex hierarchical system on the basis of data similar to the phonetic input of children.

Research in connectionism shows that network learning algorithms such as back-propagation learning in feedforward or recurrent networks are capable of learning complex mappings between representations in phonological domains (e.g. Sejnowski and Rosenberg, 1987; Weijters and Hoppenbrouwers, 1990; Gasser et al., 1989), although other experiments show that complex pattern matching problems like the extraction of syllable structure may not be an appropriate task for these methods (Daelemans and van den Bosch, forthcoming). In any case we do not have information on how the algorithm achieves the mapping and which (if any) structures are implicitly used. We propose to investigate this in a number of different ways:

- Task Analysis. By analyzing the features of the task, the interaction of learning and representation principles can be studied at a theoretical level and applied to the connectionist learning system.
- Error Analysis. Quantitative and qualitative analysis of errors on groups of forms.
- Controlled experimentation. Identify and manipulate phonological variables systematically, as in psycholinguistic research.
- Analyse activation patterns of hidden units, or connection strengths relative to categories of input data.

Furthermore, we will explore self-organizing neural networks and selectionist approaches to self-organization of phonological rules and representations, and compare their results to connectionist approaches. Daelemans (P:1991) describes an approach in which a hierarchy of phonological (assimilation) rules is learned by using a selectionist learning technique. A corpus of linguistic data constitutes the environment to a population of linguistic rules that recombine according to fitness (operationalised as the proportion of the 'environment' the rule correctly describes), and using the

idealised genetic operators crossover and mutation. In current research, a variant of this approach, genetic programming (Koza, 1991) is applied to the same and other phonological problems. The advantage of the latter approach is that the selectionist algorithm works on symbolic structures that can be inspected, interpreted and reused.

As training material and input data for these experiments, we have available on-line longitudinal observational data transcribed in UNIBET-Dutch phonematic format (Gillis, 1991), and a monomorphematic lexicon for Dutch with all formally simplex words and their phonological properties (De Schutter and Gillis, 1990).

4.8.4 Evolution of innate syntactic architectures

Recent language acquisition research within the framework of "principles and parameters" has been concerned with the question which properties of language cause a learning system to fix the parameters values. In Clark (1990, 1991), an approach based on genetic search is used to set parameters. We will investigate how this approach to learning interacts with performance aspects (parsing and generation). In work by Kempen and Vosse (1989), a model of language performance based on the metaphor of molecular interaction

The Unification Space is a hybrid architecture for syntactic processing. In this model, the lexicon and grammar are represented symbolically, while processing is based on simulated annealing, a probabilistic method. Currently, many aspects of the model are preprogrammed. We will be investigating which parts can be learned and methods for doing this will be proposed. We are considering the following specific topics:

1. **Features**. The Unification Space is a unification-based model, where feature matching partly determines the chance of unification. Currently all features necessary for unification are specified in the grammar and lexicon. It could be investigated if these can be learned on the basis of grammatical utterances.
2. **Activation levels**. The Unification Space is also an activation-based model, where the activation level of syntactic units partly determines the chance of unification. Currently the initial activation of each segment entering the Unification Space is preprogrammed. It could be investigated if the activations of specific syntactic segments can be determined by learning to account for biases in human syntactic processing.

We will also consider how the innate architectures presupposed in the above-mentioned architectures may be "evolved" themselves. In other words, how can "principles" be automatically extracted from primary linguistic data.

4.8.5 Acquisition of lexical knowledge from corpora

The main bottleneck in the design of practical, portable NLP systems is the construction of lexical databases, especially in systems that are *lexicalist*, which seems to have become the de facto standard recently in systems based on unification-based grammar theories. These systems delegate a large amount of syntactic sophistication from the grammar component to the lexicon. At first sight, this seems to make the task of automatically extracting the necessary lexical knowledge even more formidable, but this shift in computational linguistic practice also creates new opportunities for the application of self-organization techniques.

A large amount of research has been done recently on the design and organization of lexical databases in a lexicalist context. One concept that has emerged from different strands of this research is that of hierachically organized systems of categories (also called classes, prototypes, objects, nodes, typed feature structures) and also the concept of *inheritance* for minimal redundancy by sharing structure (see Daelemans, De Smedt and Gazdar, 1992 for an overview, and Daelemans and Gazdar, eds. 1991 for a number of examples).

Earlier approaches to lexical acquisition have implicitly or explicitly (e.g. in the Acquilex Project) adopted this view of the lexicon, but have been limited to symbolic pattern matching approaches to extracting taxonomies of content words based on meaning definitions (e.g. Alshawi, 1989 in the context of Acquilex).

We share the goals of Michael Brent (1991) who tries to automatically learn syntactic and semantic properties of lexical entries through unsupervised learning. In addition we want to investigate how self-organization and genetic programming can be applied to arrive at a minimally redundant hierarchical structure of lexical categories. These categories then bundle a number of properties common to a number of lexical entries. Our prime target properties will be subcategorization features and selection restrictions because they provide us with additional information on the acquisition of ontological categories, and the linking of concepts to words.

4.8.6 Data level parallelism

When complex dynamics systems are implemented on contemporary sequential computers, the computation is prohibitively slow for the kind of applications in natural language processing envisioned in this project. Since single-processor computers are also rapidly approaching their physical limits in terms of achievable speed-up, it is necessary to explore alternatives for implementing complex dynamics systems.

From these alternatives, multiprocessor systems currently hold the most promise. These systems, consisting of a number of interconnected sequential processors, are increasingly becoming available and affordable, and offer speed-ups only limited by the number of processors.

Although multiprocessor systems with a large number of processors potentially offer impressive speed-ups, it is not trivial to fully exploit their computational power. Basically, there are two approaches to programming these systems.

The first approach, *control level parallelism*, tries to identify sequences of instructions that can operate independently of each other and thus on different processors. The primary problems with this approach are synchronizing and scheduling all processes and protecting shared resources. Because of these difficulties, the peak of performance for control level parallelism lies at the order of 10 parallel processes.

The second approach, *data level parallelism*, tries to identify large numbers of maximally independent data elements and assigns one processor to each of these data elements. The number of processors that can be used effectively is only limited by the amount of independent data elements. Therefore, data level parallelism is most appropriate for programming *massively parallel* computers.

Complex dynamics systems consist of a large collection of interconnected units, and a set of computation rules which determine how the units will evolve (Steels, 1988). These computational rules are local and are applied to all units simultaneously. This means that complex dynamics systems are inherently suited for implementation on a massively parallel computer using a data level parallel programming style.

Concretely, the goals set forward in this part of the project are:

- To formulate the complex dynamics systems used in the project as data level parallel algorithms. These algorithms should have a minimal amount of global control and global interactions between data elements in order to optimize the efficiency of a massively parallel implementation.
- To implement these data level parallel algorithms on the DAP 510, a massively parallel computer with 1,024 processors connected in a 2-D mesh. The DAP (Distributed Array of Processors), produced by Active Memory Technology Ltd in the U.K., has the same capabilities as Thinking Machine's Connection Machine. The most important difference is that the DAP has "only" 1,024 processors (a version with 4,096 processors exists) and the Connection Machine has 65,536 processors. The programming methodology however is almost identical, which gives the results of this project a potentially broader impact.
- To develop tools which will aid in implementing the applications and in collecting and interpreting the results. More specifically, a graphical user interface and monitoring system will be developed which will make it possible, *inter alia*, to inspect units of the complex dynamics system, to observe macroscopic emergent phenomena, to alter parameters, to look at graphical representations of results and to modify initial and final states and conditions.

5 References

Baker, C. 1979. Syntactic theory and the projection problem. Linguistic Inquiry 10: 533-581.

Bates, E. and MacWhinney, B. 1982. Functionalist approaches to grammar. In: Wanner, E. & Gleitman, L. eds. Language acquisition: The state of the art. Cambridge: Cambridge University Press.

Bates, E. and MacWhinney, B. 1987. Competition, variation and language learning. In MacWhinney, B. ed. Mechanisms of language acquisition. Hillsdale: Erlbaum. 157-193.

Bates, E., Bretherton, I. and Snyder, L. 1988. From first words to grammar. Individual differences and dissociable mechanisms. Cambridge: Cambridge University Press.

Bowerman, M. 1983. How do children avoid constructing an overly general grammar in the absence of feedback about what is not a sentence? Papers and Reports on Child Language Development 22:

Bowerman, M. 1987. The 'no negative evidence' problem: How do children avoid constructing an overly general general grammar? In Hawkins, J. ed. Explaining language universals. Oxford: Basil Blackwell.

Bowerman, M. 1989. Learning a semantic system: What role do cognitive predispositions play? In Rice, M. and Schiefelbusch, R. eds. The teachability of language. Baltimore: Paul H. Brookes.

Brent, Michael: Automatic Semantic Classification of Verbs, pp22-26 in Powers, David M W and Reeker, Larry (eds): *Machine Learning of Natural Language and Ontology*, Proc. AAAI Spring Symposium, Document D91-09, DFKI Kaiserslautern FRG, 1991.

Chomsky, N. 1965. Aspects of the theory of syntax. Cambridge: MIT Press.

Chomsky, N. 1981. Principles and parameters in syntactic theory. In Hornstein, N. and Lightfoot, D. eds. Explanations in linguistics. London: Longman.

Clark, R. "On the Relationship between Input Data and Parameter Setting" *Proceedings of the nineteenth meeting of the North Eastern Linguistics Society*, 1988.

Clark, R. "Causality and Parameter Setting" *Behavioral and Brain Sciences*, June, 1989.

Clark, R. "Incremental Acquisition and a Parameterized Model of Grammar" in W. Sieg (ed) *Reflecting and Acting*, Kluwer, 1990.

Clark, R. "Learning from Ambiguous Data," SGAICO 90, Genève, September 1990.

Clark, R. *Papers on Learnability and Natural Selection*, Technical Report in Formal and Computational Linguistics, No. 1, Département de Linguistique, Université de Genève, 1990.

Clark, R. and Roberts I., "A Computational Model of Language Learnability and Language Change", Technical Report in Formal and Computational Linguistics, No. 3, Départemente de Linguistique, Université de Genève, 1991.

Clark, R. "A Computational Model of Parameter Setting", pp32-37 in Powers, David M W and Reeker Larry (eds): *Machine Learning of Natural Language and Ontology*, Proc. AAAI Spring Symposium, DFKI Document D91-09, Kaiserslautern FRG.

Daelemans, W. 'A Tool for the Automatic Creation, Extension and Updating of Lexical Knowledge Bases.' Proceedings of the Third ACL European Chapter Conference, Copenhagen, p. 70-74, 1987.

Daelemans, W. Studies in Language Technology: An Object-Oriented Computer Model of Morpho-phonological Aspects of Dutch. Unpublished Doctoral Dissertation, University of Leuven, 1987.

Daelemans, W. 'A Model of Dutch Morphophonology and its Applications.' AI Communications, Volume 1 number 2, p. 18-25, 1988. (Also available as AI-LAB Brussels AI-MEMO 88-8).

Daelemans, W. 'Object-oriented Hierarchical Lexicons.' Informal Proceedings of the IJCAI Workshop on Object-oriented Programming in AI, Detroit, MI, 1989.

Daelemans, W. 'Inheritance in Object-oriented Natural Language Processing.' In: Daelemans, W. and G. Gazdar (eds.) Inheritance in Natural Language Processing: Workshop Proceedings, Tilburg: ITK, ITK Proceedings 90/1, 1990, p.30-38.

Daelemans, W. and G. Gazdar (eds.) Inheritance in Natural Language Processing: Workshop Proceedings Tilburg: ITK, ITK Proceedings 90/1, 1990, 120p.

Daelemans, W. 'Biologische metaforen in natuurlijke taal verwerking' (Biological Metaphors in Natural Language Processing) Proceedings Colloquium Cultuurwetenschappen in beweging, 1991, to appear.

Daelemans, W. and G. Gazdar (guest eds.) Computational Linguistics Special Issue on Inheritance in natural language processing. Vol. 18 (1), 1992.

Daelemans, W, K. De Smedt and G. Gazdar. 'Inheritance in Natural Language Processing' Computational Linguistics, 18 (1), 1992.

Dengel A. and Barth G. High Level Document Analysis Guided by Geometric Aspects, Intl Jnl on Pattern Recognition and AI, Vol 2:4, pp641-656, 1988.

Dengel, A. and Schweizer, E. Rotationswinkelbestimmung in binaeren Dokumentenbildern (Determination of rotational angle in binary document images), in H. Burkardt et al. (eds) Proc. 11. DAGM Symposium, pp.274-278, Hamburg Oct 1989, Springer-Verlag.

Dresher, B. E. and J. D. Kaye, "A computational learning model for metrical phonology," Cognition, vol. 34, pp. 137-195, 1990.

Emde, W.: An Inference Engine for Representing Multiple Theories In: Knowledge Representation and Organization in Machine Learning, K. Morik (ed.), Springer, Lecture Notes, 1989 148–176

Emde, W.: Managing Lexical Knowledge in LEU/2 In: Text-understanding in LILOG: Integrating Computational Linguistics and Artificial Intelligence, O. Herzog / C.-R. Rollinger (eds.), Springer, Lecture Notes, (to appear October 1991)

Gillis, S. 'On prelinguistic prerequisites for linguistic reference'. In: Daalder, S. & Gerritsen, M. (eds.) Linguistics in the Netherlands 1982. Amsterdam: North-Holland. pp. 157-165.

Gillis, S. and A.M. Schaerlaekens (eds.) *Pragmatic Aspects of Early Lexical Acquisition and Development.* (Theme issue of Journal of Pragmatics) Amsterdam: North-Holland, 1985.

Gillis, S. 'Towards the first word'. In: Spoelders, M., Van Besien, F., Lowenthal, F. & Van Damme, F. (eds.) Language acquisition and learning: Essays in developmental pragmatics. Leuven: Acco, 1985. pp. 41-49.

Gillis, S. 'The development of the concept of reference in its interactional context'. Studia Filologiczne 21: 73-89, 1985.

Gillis, S. *This Child's Nominal Insight is Actually a Process: the Plateau-stage and the Vocabulary Spurt in Early Lexical Development.* Antwerp Papers in Linguistics 45, 1986.72 p.

Gillis, S. and A. Verlinden. *Nouns and Verbs in Early Lexical Development.* Antwerp Papers in Linguistics 54, 1988. 85 p.

Gillis, S., F. Koopmans-van Beinum and J. van der Stelt 'The prelinguistic stage'. In: Jordens, P. & Lalleman, J. (eds.) *Language development.* Dordrecht: Foris. 1988. pp. 29-56.

Gillis, S. *What's the Use of User Models? Linguistic Pragmatics and AI.* Antwerp Papers in Linguistics 61, 1989.

Gillis, S. *Structurele Aspecten van het Nederlandse Lexicon.* Antwerp Papers In Linguistics 64, 1990.

Gillis, S. and G. De Schutter. 'Transitional phenomena revisited: Insights into the nominal insight'. In: Lindblom, B. & Zetterstrm, R. (eds.) *Precursors of early speech.* New York: Stockton Press, 1986. pp. 127-142.

Gold E. M., "Language Identification in the Limit", Information and Control 10:447-474, 1967

Grimshaw, J. 1981. Form, function and the language acquisition device. In: Baker, C. and McCarthy, J. eds. The logical problem of language acquisition. Cambridge: MIT Press.

Hardy-Brown, K. 1983. Universals and individual differences: Disentangling two approaches to the study of language acquisition. Developmental Psychology 19: 610-624.

Harnad, Stevan (Ed.), Categorical perception: The groundwork of Cognition, Cambridge University Press, New York NY, 1987.

Harnad, Stevan, Hanson, Stephen J and Lubin J: Categorical Perception and the Evolution of Supervised Learning in Neural Nets, pp65-74 in Powers, David M W and Reeker, Larry (eds): *Machine Learning of Natural Language and Ontology,* Proc. AAAI Spring Symposium, Document D91-09, DFKI Kaiserslautern FRG, 1991.

Harnad, Stevan, "Metaphor and mental duality," in Language, mind and brain, ed. T. Simon and R. Scholes, Lawrence Erlbaum, Hillsdale NJ, 1982.

Hirsh-Pasek, K., Kemler Nelson, D., Jusczyk, P., Cassidy, K., Druss, B. and Kennedy, L. 1987. Clauses as perceptual units for young infants. Cognition 26: 269-286.

De Houwer, A. and Gillis, S. 'Language and communication during the second year of life'. Journal of Pragmatics 9: 683-704, 1985.

Huey, E. B., The Psychology and Pedagogy of Reading, MIT Press, Cambridge MA, 1908/1968.

Hume, David, Creating Interactive Worlds with Multiple Actors, B.Sc. Honours Thesis, Electrical Engineering and Computer Science, University of New South Wales, Sydney, AUSTRALIA, November 1984. (Supervisors: David M. W. Powers, Graham B. McMahon.).

Kemler Nelson, D., Hirsh-Pasek, K., Jusczyk, P., Cassidy, K. 1989. How prosodic cues in motherese might assist language learning. Journal of Child Language 16: 55-68.

Kempen, G. & Vosse, Th. 1989. Incremental syntactic tree formation in human sentence processing: a cognitive architecture based on activation decay and simulation annealing. Connection Science 1, 275-292.

Lakoff, G. 1987. Women, Fire, and Dangerous things. Chicago: University of Chicago Press.

Lakoff, George and Mark Johnson, Metaphors we Live By, University of Chicago Press, 1980.

MacWhinney, B. 1987. The competition model. In MacWhinney, B. ed. Mechanisms of language acquisition. Hillsdale: Erlbaum. 249-308.

MacWhinney, B. and Bates, E. eds. 1990. The cross-linguistic study of language processing. New York: Cambridge University Press.

MacWhinney, B. and Leinbach. J. 1991. Implementations are not conceptualizations. Ms.

MacWhinney, B. ed. Mechanisms of language acquisition. Hillsdale: Erlbaum.

MacWhinney, B., Leinbach, J., Taraban, R. and McDonald, J. 1989. Language learning: Cues or rules? Journal of Memory and Language 28: 255-277.

MacWhinney, B., Mechanisms of Language Acquisition, Lawrence Erlbaum Associates, Hillsdale NJ, 1986.

MacWhinney, B., "Conditions on acquisitional models", Proc. Annual Conference of the ACM, New York, 1978.

Macnamara, J. 1982. Names for things. Cambridge: MIT Press.

Magerman, David: Mutual Information, Deducing Linguistic Structure, pp122a-e in Powers, David M W and Reeker, Larry (eds): *Machine Learning of Natural Language and Ontology*, Proc. AAAI Spring Symposium, Document D91-09, DFKI Kaiserslautern FRG, 1991.

Maratsos, M. 1982. The child's construction of grammatical categories. In Wanner, E. and Gleitman, L. eds. Language acquisition. New York: Cambridge U.P.

Maratsos, M. 1983. Some current issues in the study of the acquisition of grammar. In Mussen, P. ed. Carmichael's manual of child psychology. 4th ed. New York: Wiley.

Maratsos, M. 1990. Are actions to verbs as objects are to nouns? On the differential semantic bases of form, class, category. Linguistics 28: 1351- 1379.

Maratsos, M. and Chalkley, M. 1981. The internal language of children's syntax: the ontogenesis and representation of syntactic categories. In Nelson, K.E. ed. Children's language. Vol. 2. New York: Gardner.

Marcus, M., A Theory of Syntactic Recognition for Natural Language, MIT Press, Cambridge MA, 1980.

Marcus, Mitchell: The Automatic Acquisition of Linguistic Structure from Large Corpora: An Overview of Work at the University of Pennsylvania, pp123-126 in Powers, David M W and Reeker, Larry (eds): *Machine Learning of Natural Language and Ontology*, Proc. AAAI Spring Symposium, Document D91-09, DFKI Kaiserslautern FRG, 1991.

Mehler, J., P. Jusczyk, G. Lambertz, N. Halsted, J. Bertoncini, and C. Amiel-Tison, "A precursor of language acquisition in young infants," Cognition, vol. 29, pp. 143- 176, 1988.

Miller, George A. and Susan M. Ervin, "The Development of Grammar in Child Language," in Acquisition of Language, ed. U. Bellugi and R. Brown, Society for Research in Child Development Monographs, 1965.

Miller, George A., The Psychology of Communication, Allen Lane: Penguin Press, London, 1967.

Miller, George A., "The magical number seven, plus or minus two: Some limits on our capacity for processing information.," Psychol. Rev., vol. 63, pp. 81-97, 1956. Republished in George A. Miller, 'The Psychology of Communication.', 1967.

Miller, P. L, "An Adaptive Natural Language System that Listens, Asks and Learns," Proc. 4th IJCAI, 1975.

Morgan, J. and Newport, E. 1981. The role of constituent structure in the induction of an artificial language. Journal of Verbal Learning and Verbal Behavior 20: 67-85.

Nicolis, John S. (1991a) "Chaos and Information Processing", World Scientific, Singapore/London, 1991.

Nicolis, John S. (1991b) "Chaotic Dynamics of Linguistic Processes at the syntactical, semantic and Pragmatic Levels: Zipf's law, Pattern Recognition and Decision Making under Uncertainty and Conflict", Proceedings of QUALICO'91, University of Trier, September 23-27, 1991.

Pike, Kenneth L. and E. G. Pike, Grammatical Analysis, Summer Institute of Linguistics (and University of Texas at Arlington), Dallas, Texas, 1977.

Pike, Kenneth L., Language in Relation to a Unified Theory of the Structure of Human Behavior, Mouton, The Hague, Holland, 1954/1967.

Pike, Kenneth L., Phonemics, Summer Institute of Linguistics, Santa Ana CA, 1947.

Pinker, 1985. Language learnability and children's language: a multi-faceted approach. In Nelson, K.E. ed. Children's language. Vol. 5. Hillsdale: Erlbaum.

Pinker, S 1982. A theory of the acquisition of lexical interpretive grammars. In Bresnan, J. ed. The mental representation of grammatical relations. Cambridge: MIT Press.

Pinker, S. 1982. A theory of the acquisition of lexical interpretive grammars. In Bresnan, J. ed. The mental representation of grammatical relations. Cambridge: MIT Press.

Pinker, S. 1984. Language learnability and language development. Cambridge: Harvard University Press.

Pinker, S. 1985. Language learnability and children's language: a multi- faceted approach. In Nelson, K.E. ed. Children's language. Vol. 5. Hillsdale: Erlbaum.

Pinker, S. 1987. The bootstrapping problem in language acquisition. In MacWhinney, B. ed. Mechanisms of language acquisition. Hillsdale: Erlbaum.

Pinker, S. 1989a. Learnability and cognition. Cambridge: MIT Press.

Pinker, S. 1989b. Resolving a learnability paradox in the acquisition of the verb lexicon. In Rice, M. and Schiefelbusch, R. eds. The teachability of language. Baltimore: Paul H. Brookes.

Pinker, Steven and A. Prince, "On language and connectionism: Analysis of a parallel distributed processing model of language acquisition," Cognition, vol. 3, pp. 73-193, 1988.

Pinker, Steven, "Concept, Word, and Sentence: Interrelations in Acquisition and Development," Cognition, vol. 7, pp. 217-283, 1979.

Plunkett, K. and Marchman, V. 1991. U-shaped learning and frequency effects in a multi-layered perceptron. Cognition 38: 43-102.

David M. W. Powers, "Neurolinguistics and Psycholinguistics as a Basis for Computer Acquisition of Natural Language," SIGART(84), pp. 29-34 (June 1983).

David M. W. Powers, "Experiments in Computer Learning of Natural Language", Proc. Aust. Comp. Conf., Sydney NSW, pp. 489-500 (Nov 1984).

David M. W. Powers and Christopher Turk, "Machine Learning of Natural Language", Research Monograph, Springer-Verlag, 1989.

David M. W. Powers, "Goals, Issues and Directions in Machine Learning of Natural Language and Ontology." Published in Powers(1991c) 1991a.

David M. W. Powers, "How far can self-organization go? Results in unsupervised language learning." Published in Powers(1991c) 1991b.

David M. W. Powers and Larry Reeker, eds., "Machine Learning of Natural Language and Ontology", Proceedings of the AAAI Spring Symposium, DFKI Document D91-09, Kaiserslautern FRG, 1991c.

David M. W. Powers, Larry Reeker and Bernhard Humm, eds., "Natural Language Learning", *Proceedings of IJCAI-91 Workshop W-14*, 1991g.

Powers, David M. W. (1991e) On the promise of learning and the problems of modularity: A biased view of Natural Language Learning, pp63-70 in Powers, David M W, Reeker, Larry and Humm, Bernhard (eds): *Natural Language Learning*, Proc. IJCAI'91 Workshop W-14, Sydney NSW 1991.

David M. W. Powers (1992a) "Autonomy, Modularity and Necessity." Contribution to Delancey & Goldsmith, *Autonomy*, to be published by Chicago University Press, 1992.

David M. W. Powers (1992b) "Learning about Learning", column to be published, SIGART 3:1, 1992.

Rieger, Burghard, B. (1989) "Unscharfe Semantik. Die empirische Analyse, quantitative Beschreibung, formale Repraesentation und prozedurale Modellierung vager Wortbedeutungen in Texten. Lang: Bern/Frankfurt, 1989.

Rieger, Burghard, B. (1977) "Theorie der unscharfen Mengen und empirische Textanalyse" (Theory of fuzzy sets and empircal text analysis) in: Klein, W. (ed) Methoden der Textanalyse, Quelle and Meyer: Heidelberg, 1977.

Rieger, Burghard, B. (1979) "Fuzzy Structural Semantics. On a generative model of vague natural language meaning" in: Trappl, R., Hanika, P., Pichler, F.R. (eds) Progress in Cybernetics and Systems Research (Vol. V), pp 495-503, Wiley: Washington/London, 1979.

Rieger, Burghard, B. (1983) "Clusters in Semantic Space" in Delatte, L. (Ed) Actes du Congres International Informatique et Sciences Humaines, pp 805-814, Laboratoire d'Analyse Statstique des Langues Anciennes: Lieges, 1983.

Rieger, Burghard, B. (1990) "Wissenserwerb aus natuerlichsprachlichen Texten. Zur Interdisziplinaritaet eines computerlinguistischen Forschungsprogramms" (Knowledge Acquisition from natural language texts. On the interdisciplinarity of a computational linguistics research program.) in: Kuehlwein, W., Raasch, A. (Eds) Angewandte Linguistik heute pp101-124, Lang: Frankfurt/New York, 1990

Rumelhart, D. and McClelland, J. 1986. On learning the past tenses of English verbs. In McClelland, J. and Rumelhart, D. eds. Parallel distributed processing. Vol. 2: Psychological and biological models. Cambridge: MIT Press. 216-271.

Samuel, A. L., "Some studies in machine learning using the game of checkers II recent progress," IBM Jnl of R & D, vol. 11, no. 6, pp. 601-617, 1967.

Scha, R. (1990). Taaltheorie en Taaltechnologie; Competence en Performence (In Dutch). In: Computertoepassingen in de Neerlandistiek (K. de Kort & G.L.J. Leerdam, Eds.), pp. 99-109. Almere: LVVN.

Schaerlaekens, A. and Gillis, S. *De taalverwerving van het kind.* Groningen: Wolters-Noordhoff, 1987

Scholtes, J.C. (1990). Trends in Neurolinguistics. Proceedings of the IEEE Symposium on Neural Networks, 21st June 1990, Delft, Netherlands, pp. 69-86.

Scholtes, J.C. (1991). Using Extended Kohonen-Feature Maps in a Language Acquisition Model. Proceedings of the 2nd Australian Conference on Neural Networks, Sydney, February 2nd-4th 1991.

Scholtes, J.C. (1991). Learning Simple Semantics by Self- Organization. Proceedings of the AAAI Symposium on Connectionist Learning of Natural Language, Palo Alto, 26st-29th March 1991.

Scholtes, J.C. (1991). Learning Simple Semantics by Self- Organization. Proceedings of the AAAI Symposium on Machine Learning of Natural Language and Ontology, Palo Alto, 26st-29th March 1991.

Scholtes, J.C. (1991). Recurrent Kohonen Self-Organization in Natural Language Processing. Proceedings of the ICANN, pp. 1751-1754. June 26th- 29th 1991, Helsinki Finland.

Scholtes, J.C. (1991). Unsupervised Context Learning in Natural Language Processing. Proceedings of the IJCNN, Vol. 1, pp. 107-112. July 8th-12th 1991, Seattle.

Scholtes, J.C. (1991). Kohonen Feature Maps in Natural Language Processing. Technical Report, March 1991, Department of Computational Linguistics, University of Amsterdam.

Scholtes, J.C. (1991). Unsupervised Learning and the Information Retrieval Problem. Proceedings of the the IJCNN, November 18-21, Singapore.

Scholtes, J.C. (1991). Neural Nets and Their Relevance for Information Retrieval. Technical Report, August 1991, Department of Computational Linguistics, University of Amsterdam.

Scholtes, J.C. (1992). Neural Nets versus Statistics in Information Retrieval. A Case Study of the 1987 Pravda. Proceedings of the SPIE Conference on Applications of Artificial Neural Networks III. 20-24 April 1992. Orlando, Florida.

Sejnowski, T. and Rosenberg, C. 1987. Parallel networks that learn to pronounce English text. Complex Systems 1: 145-168.

Slobin, D.I. 1985. Cross-linguistic evidence for the language-making capacity. In Slobin, D.I. ed. The cross-linguistic study of language acquisition. Vol. 2. Hillsdale: Erlbaum.

Slobin, Dan I., The Ontogenesis of Language, Academic Press, New York, 1971.

Slobin, Dan I., "Imitation and Grammatical Development in Children," in Contemporary Issues in Developmental Psychology, ed. E. Endler, L. Boulter, and H. Osser, pp. 437-443, Holt, Rhinehart and Winston, New York, 1968.

Sparck-Jones, K. "Notes and references on early automatic classification work", unpublished notes.

Sparck-Jones, K. "Synonymy and semantic classification" Edinburgh UP: Edinburgh UK, 1986 (publication of her early thesis work of 1964).

Sparck-Jones, Karen, "Mechanized Semantic Classification," 1961 Conference on Machine Translation and Applied Language Analysis, vol. II, pp. 418-435, Her Majesty's Stationery Office, London, 1961.

Spiessens, P. (1990), "PCS: A Classifier System that Builds a Predictive Internal World Model", In: Proceedings of the European Conference on Artificial Intelligence (ECAI-90), Pitman Publishing, London.

Steels, L. (1988) "Artificial Intelligence and Complex Dynamics" in Concepts and Techniques of Knowledge-based Systems, Tokoro M. (ed).

Steels, L. (1988) "Steps Towards Common Sense" in Proceedings of the 8th European Conference on Artificial Intelligence (ECAI-88).

G. Tonfoni, La Lettura Strategica: Tecniche Cognitive per Leggere di Piu e Meglio (Strategic Reading: Reading Techniques to Get More Out of the Text), with Tassi, Mondadori Informatics, 1990. (A second edition of this book is to be published during the summer of 1991.)

G. Tonfoni, Didattica del Testo: Curricolo di Formazione Linguistica per gli Insegnanti (Teaching through Texts: The Teacher's Guide to Text-Writing and Reading Class Experiences), Lisciani e Giunti, 1991.

Turk, C. C. R., "Evolutionary Methodology in the Acquisition of Natural Language by Machine Learning" in Trappl, R. (ed.) *Cybernetics and Systems '88*, Proc. of the Ninth European Meeting, Riedel: Amsterdam, 1988 (copresented by David Powers).

Turk, C. C. R., "Evolutionary Methodology in Natural Language Acquisition by Machine Learning" pp887-894 in Trappl, R. (ed.) *Cybernetics and Systems '90*,

Proc. of the Tenth European Meeting, World Scientific: Singapore/London, 1990 (copresented by David Powers).

Verlinden, A. and Gillis, S. 1988. Nouns and verbs in the input: Gentner (1982) reconsidered. In: Van Besien, F. (ed.) First language acquisition. ABLA Papers 12: 163-187.

Wanner, E. and Gleitman, L. eds. Language acquisition: The state of the art. Cambridge: Cambridge University Press.

Weijters, T. and Hoppenbrouwers, G. 1990. Netspraak: Een neuraal netwerk voor grafeem-foneem-omzetting. Tabu 20: 3-28.

Wijnen, F. 1990. On the development of language production mechanisms. PhD, KU Nijmegen.

Winston, P. H., "Learning structural descriptions from examples," in The Psychology of Computer Vision, McGraw-Hill, 1975.

Winston, P. H., *Artificial Intelligence*, Addison Wesley, 1983.

Bibliotheek K. U. Brabant
17 000 01173168 5

ITK: P.O. BOX 90153 5000 LE TILBURG THE NETHERLANDS