**Bibliographies and Literature Reviews**

**Goals, Issues and Directions in Machine Learning of Natural Language and Ontology**

David M. W. Powers,
FB Informatik University of Kaiserslautern, FRG
powers@informatik uni-kl.de

1. Introduction

Over the last thirty years there has been a trickle of papers addressing aspects of the Natural Language Learning area. The 80s have even seen a few books published on the subject. These have tended to take drastically different theoretical approaches, and have drawn on varying degrees on fields outside Computer Science and Artificial Intelligence.

During this same period, computational and mathematical modelling of language and learning have increasingly been recognized as relevant to assessing the validity of a theory of Language Acquisition or the Nature of Language. Conversely, researchers in Linguistics, Psycholinguistics and Philosophy, as well as Computer Science, have been considering how and where we can apply our increasing knowledge of the human characteristics and constraints which determine how we solve problems, learn about the world, and use language.

This review was originally prepared as a background paper for the AAAI Spring Symposium on Machine Learning of Natural Language and Ontology to be held in Stanford (March 26–28, 1991), which will address all aspects of the relationship between Machine Learning and Natural Language. We not only expect input from researchers in Computer Science and Artificial Intelligence (Machine Learning, Natural Language, Robotics, Vision, Neural Nets, Parallelism, etc.) but wish particularly to encourage relevant contributions from other fields (Linguistics, Psycholinguistics, Philosophy, Neurology, Mathematics, etc.)

The following seeks to expand on the perceived extent of the field, to clarify and encourage the possibility for participation from those with potentially relevant research, and to indicate some of the areas in which research is, can be or should be focussed. Indications are that the time is becoming ripe for a major effort on Language Learning. Researchers who have an interest in the topic, including those who have not been active in Machine Learning of Natural Language but related fields, are encouraged to contact the author and indicate their interest and work.

Our descriptions are intended to be totally independent of application, whether Database Retrieval, Machine Translation, Modelling (Linguistic, Psychological or Neurological), or Robotic Speech and Vision systems.

The symposium will address all aspects of the relationship between Machine Learning and Natural Language. We not only expect input from researchers in Computer Science and Artificial Intelligence (Machine Learning, Natural Language, Robotics, Vision, Neural Nets, Parallelism, etc.) but wish particularly to encourage relevant contributions from other fields (Linguistics, Psycholinguistics, Philosophy, Neurology, Mathematics, etc.). This paper is similarly broad.

Specific areas of interest, selected according to the author's own sense of the field, will now be reviewed. Representative or exhaustive references will be included under each major heading with some annotation. Notification of errors and significant omissions will be welcomed. The purpose of this document is to stimulate interaction and expand our perceptions of the field.

Note that the last of the six major headings, System Development, represents the practical outworking of our field and aims to be exhaustive (at least of the researchers and groups, if not of their output), and where possible comparative. The preceding sections refer to work which has proven useful, or has been proposed as useful, in relation to this ultimate aim of building an effective system. Emphasis is on the roots of these other fields rather than the modern breadth which they encompass.

2. Traditional Approaches

The first broad area which we are interested in presenting is a mix of disciplines – although in the foundational work presented here the borders are already becoming fuzzy. We are interested in the basic work in Linguistics and Machine Learning which might be expected to be relevant to MLNL research, and in most cases has in fact formed the basis of some research projects in this area. Psycholinguistics and Nerolinguistics, however, are treated under the modern banner of "Cognitive Science" in the following section.

The traditional natural language work has been based in varying degrees on linguistic theories and models. Machine learning has largely been focussed on concept learning, heuristic evaluation with signature tables, information theoretic discrimination learning, etc. This bibliography summarizes the roots of both natural language and machine learning work. The many modern texts are not referenced, but could be referred to for up-to-date detail.

2.1. Applicability of traditional machine learning

2.1.1. Introduction

Under the heading of Machine Learning, we particularly have in mind work in concept learning – clearly related to semantics and potentially to syntax and pragmatics. We are also interested in the role of teacher and critic, including automatic generation of examples, implicit criticism, unsupervised learning etc. Application of traditional techniques to facets of language are fundamental in that they are immediately accessible and connect with a considerable body of previous work.

2.1.2. Bibliography


interesting for its showing that the role of teacher may be separated from that of critic. In Sammut's system, after a generalization step the system provides its own new example to test the validity of the generalization, and only requires positive or negative criticism. The teacher need only provide the initial (positive) example. The critic must provide feedback on every example. This type of approach is particularly appropriate for learning of semantics. It is primarily in a neural network or statistical context that I am aware of inductive learning applications where criticism is not used (see sections 4 & 5).

But there are types of learning other than induction, the learning of new concepts or rules. There is also learning to do things better or faster. Explanation-based learning (Mitchell et al., DeJong and Mooney), the version space technique (Mitchell), EURISKO (Lenat) and Chunking (Laird et al.) have also their applications to MLNL.

We note that Lenat's more recent work on CYC, which uses explicit acquisition rather than machine learning in the present stage of the project, deals with other problems related to MLNLO and is referenced in section 6. Bedewith et al's work with Miller (see section 4) is in some ways similar, concentrating on different directions than MLNL at the moment in the application of psycholinguistic results.

A final classification of learning systems can be made on the basis of whether they are capable of incremental learning or not. Winston, Fisher, Shapiro and Riesbeck particularly address this "problem". Some of the above techniques like to work with full information, or a sample, others are inherently incremental. While some see restriction to incremental learning as a disadvantage, given that Natural Language can be learnt with incremental exposure it could well be that incremental algorithms can be more efficient for a class of problems which includes MLNL (see section 3).

2.2. Applicability of traditional linguistics and parsing techniques

2.2.1. Introduction

Some MLNL approaches are based on traditional theories from linguistics and elsewhere. Learnability provides a very practical test for a linguistic theory. A good approach to parsing should relate to a good approach to learning syntax. Many approaches however are based on non-linguistic traditions, notably neural nets. It is especially important to consider the connections between different disciplinary approaches.

2.2.2. References


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


2.2.3. Significance

The above references have been advertized as traditional linguistics and parsing techniques. Some therefore require a word of explanation on their inclusion!

The behaviourist references, to Skinner and critiques of his work, are included here because of his critiques of linguistics, not to mention the confrontation between his approach and Chomsky's (which I haven't documented here). The reference to Popper's philosophy of science doesn't really belong under Machine Learning either, but it is fundamental to some of the issues in Linguistics today, and it is also relevant to Machine Learning of Ontology -- in the sense that Science is the process by which, as a society, we learn about our world.

Some of the AI work, Schubert's for example, is quite a way from parsing, but deals with issues important to semantics, and is part of the heritage we have when we come to do MLNL. Other work, Schank's and Wood's, are particularly fundamental tradi-
...tions in NL. Schank is concerned also with semantics, and conceptual dependency theory is one of the most well developed semantic representations. The work of his group moreover extends to MLNL projects (see section 7). Pereira and Warren are particularly important as representatives of the Logic Programming approach to NLL, and learning techniques have also been applied to their work (see section 7).

Fiske is represented for his broad view of language and behaviour, being one of the first to recognize that language and ontology cannot be separated. His theory and methodology of Phonology are still standard, and his generalization to the Tagmemics theory of grammar is significant for its supporting of phrase structure with cohesion, and has also proven the base for some MLNL work (see section 7).

Halliday is responsible for the Systemics grammatical theory brought to the attention of the AI world by Winograd. He also emphasizes the role of cohesion.

But even after 30 years, Chomsky's school of Transformational Generative Grammar (TGG) remains dominant. It has, moreover, had a significant influence on Psycholinguistics (see section 4), which in turn has given rise to TGG, represented here by Derwing and Vanderslice. It too has been used as a guide for MLNL work, and TGG has benefited from criticism from this source as well (see section 7).

2.3. Goals and Issues

*Goal:* Theories of language (grammar, semantics, representation) and learning (reasoning, understanding cognition) which allow effective language learning and use.

*Issue:* Will traditional approaches to language and learning be effective in some combination? Or will language learning require/lead to new theories of language and learning?

3. Complexity Theory

3.1. Formal results on learning and language constraints

3.1.1. Introduction

Results and proposals based on complexity theory have been driving forces in some schools of linguistics and psycholinguistics—notably the contributions of Gold and Chomsky. New approaches, algorithms and claims need to be considered in the light of such results, and appropriate new analyses should be developed.

Rigorous mathematical analysis is an important source of criticism for Cognitive Science research. Publication of results can shape the whole future of a field, firmly closing off former paths of attack, and opening up others. Unfortunately, the effect has not always been positive. In some noteworthy cases, the wider Cognitive Science community has taken a result at face value, applied it far outside the applicable conditions (spelled out by the original author), and interpreted it without common sense reflection on and reinterpretation of the natural world correlates of the analyzed system. This list includes a number of such examples. It pays to consider these results first hand!

3.1.2. Bibliography


3.1.3. Significance

As indicated above, some of the work here has single-handedly changed the course of history, to negative as well as positive effect.

Minsky (with Papert) showed that there were certain classes of problem which were not learnable with certain networks of perceptions. Still today (comp.ai newsgroup 19 Oct 90) he is fighting the widespread belief that he killed perceptions but that now, finally, connectionism has laid to rest Perceptrons, the book. As he says in comp.ai: "Try reading the book." (which has recently come out in an expanded edition). Actually, he has maintained his research interest in this area over the missing years. Even today there is a need for the theoretical analyses here: the psycholinguistic evidence about how, what and how long we learn (section 3) and the connectionist wave of practical successes (section 4), need to be brought together to reconcile our expectations about what we should be able to build easily with the evidence about what we can actually do easily. Then we will be able to start building systems appropriate to the tasks.

Chomsky and Miller also set Linguistics and Psycholinguistics on a new track with their importation of formal analysis techniques. However, the accuracy of this diagnosis of linguistics does not automatically imply the uniqueness or even soundness of Chomsky's remedy. But the resulting massive persuasion to TGG (section 1.2) has generated a mass of useful research which has given this theoretical approach unprecedented (in Linguistics) opportunity for refinement.

Gold has also set a conundrum for Psycholinguistics. If Natural Language is a Context Free Language (CFL) (see Pullum and Gazdar and Postal and Langendoen) and if our parents don't provide us with the criticism necessary to learn a CFL and our environment doesn't somehow provide us input in a textbook order (see section 3 - Psycholinguists are convinced these conditions aren't satisfied) then we cannot learn Natural Languages.

Chomsky's answer was that we don't learn language, but select a subset of an innate super-language to be appropriate to our language environment. Another approach is to see if we can come up with a closer classification of Natural Language and the Psycholinguistic and Environmental restrictions.

3.2. Development of effective classifications of language

3.2.1. Introduction

Part of the problem with formal theory is the lack of evidence that the theoretical classification of language relates to the actual human languages and cognitive restrictions. Some basic assumptions are clearly suspect or at least oversimplifications. Do we need to develop new ways of formally characterizing language in terms of the restrictions and heuristics which shape human learning of language?

The lack of references under this head is indicative of a significant lacuna.

3.2.2. Bibliography


3.2.3. Significance

Miller, cited earlier for his work with Chomsky, made another important contribution: on the Magic Number Seven. This is included in the next section. Yngve was the first to put such restrictions to positive effect, implicitly restricting the class of languages he was trying to analyze.

3.3. Goals and Issues

Goal: Theoretical analysis is needed to determine and characterize the relation between supervision level, computational constraints, formal language class and base level knowledge.

Issue: The positive effect of negative constraints on the computational capacity has been neglected. Such constraints effectively define new subclasses of languages learnable by a given algorithm. The languages humans encounter are not arbitrary but are shaped by our algorithms, limitations and environmental (including supervisory) conditions, being limited to what can be learned (or, stronger still, invented) under these conditions.

4. Cognitive Science

4.1. Psychological results on language and restrictions

Psychological results on language and restrictions are seen as a major foundation for MLNL, with the hope that old and new results and critiques from Psycholinguistics will inspire those who are looking for solutions to problems, and ideas they can implement. For participation in the symposium, it is not necessary that the participant has himself worked on learning programs, but the relevance of his work to such efforts should be made clear.

4.2. Linguistic results on the nature of natural language

Similar considerations apply here. Comparative advice about linguistic theories or formalisms, with critical evaluation on the basis of computability, are basic to MLNL research. Implementers who have adopted a particular linguistic heritage are particularly asked to comment on the reasons for the choice plus the appropriateness in retrospect.

4.3. References

The emergence of Cognitive Science in the 80s as the interdisciplinary counterpart of Artificial Intelligence represents a huge increase in interest in the potential interdisciplinary contributions to understanding and modeling intelligence, learning and language. This is reflected here in only token form, allowing the reference to the older expositions which preempted the universalist approach and the debates which ensued and lead directly to the recognition of Cognitive Science. The linguistic and philosophical traditions have been to a greater or lesser extent reflected in the last section; whilst the new age neural developments are reflected in the next section to the extent that they are treated at all.

This leaves, in the main, Psycholinguistics.


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


Sloman, Aaron and Monica Croucher, "Why Robots will have Emotions," 7th International Joint Conference on Artificial Intelligence, pp. 197-202, 1981.


4.4. Significance

Having worked so hard to promote interdisciplinary connections and establish Cognitive Science, this is not the place to try to disentangle the woven threads into the distinct fields. The work and the researchers presented here are increasingly moving beyond the primary boundaries of the host discipline.

By far the majority of the references included here can, however, be considered as having a psycholinguistic orientation, a bias which is by no means independent of our focus on MLNL. And Piaget can be considered the father of Psycholinguistics. The reference here is actually the first of a dozen or more books on language and reasoning 'chez l'enfant'.

It goes without saying that the work of Anderson, Brown, and Newell and Simon is essential reading. Anderson for work on Memory, which has extended in to Language Acquisition models (section 7); Brown for the enormous contribution he and his co-workers have made to the analysis of child language and the direction of the field; and Newell and Simon for their work on problem solving which was actually fundamental to the inception of AI – it is often forgotten that the cognitive science elements were there from the beginning.

Pike's broad view of language is a mammoth effort showing incomparable insight into linguistic processes, and like Piaget worth making the effort to digest. The work of Lakoff and Johnson on metaphor is classic and is also essential reading – metaphor and metonymy are not just linguistic devices but are fundamental to the way we use language and understand the world. Language as we know it, ontology, just couldn't exist without extending our experience of particular situations to others. Metaphor and paradigm give us a model for employing contrast and similarity as a basis for generalization and application of knowledge. The work of Clark and Clark examines particular aspects of fundamental metaphorical usage in language and is again essential reading.

Along with a myriad of other fundamental work, the Clarks' papers are to be found in some key compendia. Six such volumes are included in the bibliography above. We select out some of the work of particular interest from them here.

Cofer and Musgrave (1963) includes Roger Brown and Colin Fraser on The Acquisition of Syntax.

Smith and Miller (1966) includes Jerry A. Fodor (How to Learn to Talk: Some Simple Ways), Eric H. Lenneberg (The Natural History of Language), David McNeill (Developmental Psycholinguistics) and Dan I. Slobin (Acquisition of Russian as a Native Language).

Slobin (1971) includes Martin D. S. Braine (On Two Types of Models of the Internalization of Grammars), David McNeill (The Capacity for the Ontogenesis of Grammar) and David S. Palermo (On Learning to Talk: Are Principles Derived from the Learning Laboratory Applicable?).

Moore (1973) includes Melissa Bowerman (Structural Relationships in Children's Utterances: Syntax or Semantic?), Eve V. Clark (What's in a Word? On the Child's Acquisition of Semantics in his First Language), Herbert H. Clark (Space, Time, Semantics, and the Child), Susan Ervin-Tripp (Some Strategies for the First Years) and Gary M. Olson (Developmental Changes in Memory and the Acquisition of Language) and H. Sinclair-de Zwart (Language Acquisition and Cognitive Development).

Fletcher and Garman (1979) includes contributions by Melissa Bowerman (The Acquisition of Complex Sentences), Bruce L. Derwing (Language Acquisition: Studies in First Language Development), Eve V. Clark (Building a Vocabulary: Words for Objects, Actions and Relations), William J. Baker, (Recent Research on the Acquisition of English Morphology), Robert Grieve and Robert Hoogenraad (First Words), Patrick Griffiths (Speech Acts and Early Sentences) and Michael P. Maratos (Learning How and When to Use Pronouns and Determiners).


In these cases there would be too much to address each contribution, but I the volumes are thoroughly worth a browse through, looking particularly at the papers mentioned.

As a final topic in Psycholinguistics, we mention the particular focus of imitation and correction, including the role of language play. Fraser, Bellugi and Brown have performed fundamental
studies here which have some surprising results about the relative difficulty of production, comprehension and imitation. Kuczaj and Derrick also extend the ideas of imitation and reduction to consider the child's own spontaneous paradigmatic production and self-imitation, and the parents use of imitation and expansion. Again there may be some surprises about just what parental language is most, and least, helpful.

There is one other collection of papers, Piatelli (1979), which is a must, and arises from consideration of the imitation and correction paradigm and its insufficiency. Here the protagonists of the great debate of nativism versus constructivism address each other's arguments in finding a position to be taken in reply. These deal directly with the issues that came out of the theoretical considerations of section 3 and represents the point of time, around the birth point of Cognitive Science, where the possibilities for reconciling theoretical and empirical results on language learning were just beginning to re-emerge in the face of rampant nativism.

Moving from those papers directly concerned with language learning, we come first to the generalization where a whole culture and language is incompletely learned and generates pidgins and creoles. In such a context where the learners come from a mix of language backgrounds and learn the words but not the grammar of a new "common" language, a new creole grammar emerges which is relatively independent of all of the original languages. Bickerton presents an interesting paper on this phenomenon. It would seem that it should contribute to our modelling of default preferences during language learning, and that it should be contrasted with child grammars. These challenges have yet to be taken up.

Moving further afield, we come to the famous paper of Miller on the "Magic Number Seven", which challenges us to take our known Cognitive Restrictions into account, and is really the key ingredient in finding a position to be taken in reply. These deal directly with the issues that came out of the theoretical considerations of section 3. And then there is Huye's classic on reading -- the only work from last century cited in this review. Techniques of eye motion and examining our reading behaviour can also provide insights into language behaviour, and help explain some of the behaviour of our learning programs too.

Extending from M1NL to MLNO, leads us to consider important work related to our ontology, and of course the visual modality which we feel is so dominant and which is one of the most well explored areas of cognition. Here the work of Hubel and Wiesel is again classic -- and has been the basis for some experiments on self-organization neural models, both for vision and language (Powers, section 4), whilst the work of Pylshyn is directly complementary to some Psycholinguistic studies.

4.5. Goals and Issues

**Goals:** To provide the empirical evidence for the roles of innate knowledge and specific and general learning mechanisms, as well as for environmental conditions including parents and other human supervisors and critics plus the physical laws and feedback deriving from physiological constraints.

**Issue:** How much is (necessarily) innate? From how minimal a base state can learning be effective in bootstrapping?

**Issue:** How much supervision, teaching and criticism is necessary for effective learning? To what extent can a reactive environment substitute? What cognitive constraints shape our languages?

5. Parallel Networks

5.1. Neural models of parsing and learning

There is a separate parallel symposium on "Connectionist Natural Language Processing". For this reason we are most concerned here with the advantages of neural approaches over conventional machine learning or with deep modelling of neurolinguistic processes, rather than with application of backpropagation in this or that area -- there is really just too much of an explosion in Connectionism to do justice to it here -- we provide the fundamental references but no more. But we explore in other directions. In particular, we are interested in hard neurological evidence and the associated theories.

5.2. Parallel Models of parsing and learning

Implementations on parallel hardware are also of interest, as are parallel or parallelized algorithms and theoretical contributions on the role of parallelism, backtracking etc. in language and learning processes.

The interest in parallel parsing goes back just as far as the roots of connectionism -- in fact there has been a long standing assumption that natural language parsing was inherently parallel. This debate is starting to favour the view that it is not, even without backtracking. But there is still evidence that our own brains do use at least a partly parallel process.

5.3. References

This very short list points to both the old school and the new age of associative and neural networks, as well as the only parallel language learning proposals I am aware of.


Charniak, E. and E. Santos, "A connectionist context-free parser which is not context-free, but then it is not really connectionist either," *Proc. 9th Conf. of the Cog. Sci. Soc.*, pp. 70-77, Seattle WA, July 1987.


Grossberg, S., "Adaptive Pattern Classification and Universal Recoding: I. Parallel Development and Coding of Neural Fea-


5.4. Significance

To start with modern Connectionism, Rumelhart and McClelland and their PDP group have put out three volumes, the last with a disk of sample programs. For a general review of the field see Pollack (1989), and in relation to natural language applications, see Selman (1989) – the other articles in the same issue of *AI Review* as these may also be of interest.

To go back to the roots, Hebb originated the plasticity hypotheses which is largely the basis for the PDP work, with the addition of backpropagation. Without this feedback, useful learning can still be done as shown by von der Malsburg in reproducing the visual cortical columns. Powers has applied the same technique to language between the grapheme and noun phrase levels.

Grossberg has been pursuing the properties of recurrence for well over a decade, originally as a model of memory, but more recently has come up with some interesting results in feature recognition. His recent book is not listed but presents his whole program through the papers he has presented over the years. Kohonen and Longuet-Higgins et al. also represent older schools concerned with associative memory properties. Amari and Arbib provides a good timestamp for the point just as Connectionism began to take off and emphasises the competitive and cooperative aspects which have been taking a back seat recently, but produce useful results when applied to appropriate problems.

Uhr is pretty good for its time, but rather dated now, and somewhat negative on the Perceptron/Self-Organization question. Waltz and Pollack is very aware of the neural aspects, whilst the parallelism of Jain and Sharma is totally independent of connectionism – their team model being more related to the OR–parallelism of logic programming or evolutionary learning.

5.5. Goals and Issues

Goal: Neural investigations need to determine and characterize the nature and role of the human (animal) wetware, as well as stretching the limits of neural inspired models.

Issue: What are the limits of genetic determination, boundary conditions and self-organizational determination?

Issue: Neural simulations to date tend to be passive recognizers reacting to the sensory–motor input. Does there exist some sort of active learning which is different, which is not just a feedback control system, but capable of initiating behaviour?

Issue: How does all of this relate to language? Is language just a consequence of our neural capacities in combination? Or are language specific neural level mechanisms to be found?

6. Symbol Grounding

6.1. Grounding of Natural Language Systems

Where is the border between syntax and semantics? When can a system be said to know something as opposed to just churning out a pat response? Does learning provide an answer to these old chestnuts? How far can you get with an ungrounded system?
6.2. Interaction between Modalities and Learning of Ontology

We are particularly interested in contributions in which aspects of language are learnt and used in a context or where language input and output are supplemented by (or indeed supplement) other forms of interaction between the language system and the environment in which it is embedded. The system could be a robot, simulated or actual; the environment could be provided by a vision system; or we could have a humber interface to a database, an operating system or other application.

6.3. References

The Cognitive Science and Theoretical Approaches literature is relevant background, the work listed here faces directly the question of the individual in relation to his world and his representation thereof. Explicit reference to Searle and Turing are avoided here as irrelevant, but you can’t explore far before falling over the ubiquitous Chinese Room.


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

Lenat, D. and R. V. Guha, “The world according to CYC,” ACAA–AI–300–88, MCC, 3500 West Balcones Center Drive, Austin TX.

To be published by Addison–Wesley in expanded form as “Building large knowledge based systems”.


Sloman, Aaron and Monica Croucher, “Why Robots will have Emotions,” 7th International Joint Conference on Artificial Intelligence, pp. 197–202, 1981.


6.4. Significance

The catch-phrase “Symbol=Grounding” has been popularized primarily by Harnad, who has been interested in it from a philosophical view and as an exponent of Total Turing Test versus the Turing Test (in relation to Searle’s Chinese Room). The problem is that no matter how many times you translate your text to a new “representation language” you still have a language composed of symbols and no possibility of real meaning or understanding. Where does our meaning and understanding come from?

The work referenced here is quite varied, varying from a text on how to learn a language monolingually (Brewester and Brewester), to various world modelling projects, including Lenat’s CyC project and Carbonell and Hood’s World Modelers Project. Lenat sees the problem of representation as being much more pressing than learning at this time, and his project is not supposed to start its automatic acquisition phase till 1994.

On a smaller scale, Hume and Powers have developed a Robot World modelling system for language learning work (used also by Sammut for concept learning; see also section 7), and many other projects have worked with or proposed similar schemes, including, as early examples, Block et al. and McCarthy et al. The Naive Physics Manifesto of Hayes encourages moving to the big time – toy world’s are only good for toy systems, so sometime we have to represent the world more realistically.

Others researchers have concentrated on particular problems or manifestations – this includes the work of Sloman and Croucher and that of Pylyshyn.

Moving further back to the fundamental psycholinguistic and neurolinguistic study of how we develop our ontology and semantics, we come to some classics, Lettvin et al. on frog vision, and Piaget and Lenneberg on language understanding in normal and abnormal circumstances respectively. Lerner and Marshall provide more recent perspectives. Then there is Jolles’ whimsical attempt to catalogue the whole of knowledge – on the basis of similarities which run orthogonally across all areas and levels of knowledge.

Again metaphor is an important part of the answer to symbolic grounding. Once the world has made an iconic image somewhere in our brain, we have an imperfect reflection of reality. We continue to abstract and manipulate this in a way determined by our experience, that is by similarity to what we have experienced in the past, in the same or in different modalities, and always in at least slightly different contexts. The frame problem arises when we make our concepts too small and forget that in fact we tend to present whole frames in each modality, and it is these we compare and process.

Irrespective of whether Harnad is right about characterizing this as being the major problem for Natural Language and Artificial Intelligence, Symbol Grounding is currently one of the weakest areas, and these few pointers here need to be paid more attention and to grow into resources for future MLNLO work.

6.5. Goals and Issues

False Issue: What is the difference between a Chinese room and a Chinese in a room? (Complexion??)

NOTE:
NONE = Turing = Outside room = Black Box
MIND = Searle = Inside room = Glass Box
SKIN = Harnad = Essence = Colored Box

Turing says that intelligence and thought is totally captured in language and can be totally expressed and communicated through arbitrary symbols.

Searle says that intelligence and thought is totally equivalent to mind and can’t be totally expressed and communicated through any symbols.

Harnad says that intelligence and thought is equivalent to human-ness and can’t be totally expressed and communicated through any symbols, but can be expressed and communicated through the right set of ‘symbols’ (which must include icons, the intrinsic counterpart of symbols).

Ultimate Goal: To have language used effectively by the computer for the purpose we intend.

Real Issue: When are we just translating from one language to another? When are we doing more: understanding, communicating, intending? Where does a computer derive its motivation from? Its programmer? Where do we derive our motivation from?

Toy Sub Goal: To provide a toy environment in which the above is achieved.

Real Sub Goal: To achieve this in an actual application environment.

Crucial Issue: How similar a sensory-motor environment and perceptual interface to ours is needed to allow learning of language? And what criterion do we learn to?

7. System Development

7.1. Computable hypotheses and heuristics for language learning

“Proposals of how to build a language learning system are few and far between, and will be received with interest, as will more limited argument about the significance of various hypotheses, heuristics or methodologies for language learning implementations.”

At the moment language learning work tends to work in a small way, examining how far one can get with certain techniques. Whether these techniques are implemented as AI or Psychological modelling, they should make a contribution to our understanding of language and learning.

7.2. Experimental language learning systems and their rationale

“Reports on successfully implemented language learning systems will be received with amazement! Characterizations of what can be learnt by the system, or any precursor thereof, should be included, along with explanations of the methodology used.”

Even when there are concrete ideas about how to proceed, there has been negligible funding for MLNLO research, and the implementations have been limited. But the field does seem to be ripe now for a major effort, and there is no shortage of ideas and methodologies coming from all the disciplinary vantage points considered in this review.
7.3. References

Currently all the MLNL work I know of lies somewhere between, and less than, providing full proposals or implementa-
tions, but research is progressing more or less scientifically, by small steps, as influenced by particular views.

I list here every known researcher who has developed any system which in any sense makes a claim to learn an aspect of lan-
guage. Not every individual publication is listed, but rather the most comprehensive and accessible.

Anderson, John R., “Computer Simulation of a Language Ac-

Anderson, John R., “A Theory of Language Acquisition Based
on General Learning Principles,” 7th International Joint Confer-
ence on Artificial Intelligence, pp. 97–103, 1981.

Berwick, Robert C., The Acquisition of Syntactic Knowledge,

Berwick, Robert C. and Sam Pilato, “Learning Syntax by Au-

Davidia Lazarro, David M. W. Powers, Debbie M. Meagher, and
David Menzies, Further Experiments in Computer Learning of
Natural Language, pp. 458–468, Sydney NSW Australia, Sep-
tember 1987. Second Australian Conference on Artificial Intelli-

Drescher, B. E. and J. D. Kaye, “A computational learning model

Feigenbaum, E. A., “The simulation of verbal learning behav-
or,” in Computers and Thought, ed. Feigenbaum and Feldman,

Granger, Richard, “FOUL-UP: a program that figures out
meanings of words from context,” Proc. 5th IJCAI, pp. 172–178,
1977.

Harris, Larry R., “A System for Primitive Natural Language
1977.

Kelley, K. L., “Early Syntactic Acquisition,” P-3719, Rand Cor-
poration, Santa Monica CA, November 1967.

Kucera, Henry, “The Learning of Grammar,” Perspectives in

Lamb, Sydney M., “On the Mechanization of Syntactic Anal-
ysis,” 1961 Conference on Machine Translation and Applied
Language Analysis, vol. II, pp. 674–685. Her Majesty’s Stationery

Langley, Pat, “Language Acquisition Through Error Recov-

idiosyncratic grammars,” Proc. 7th Int’l Machine Learning Confer-
ence, Austin TX, 1990.

Lehman, J. F., “Supporting Linguistic Consistency and Idiosyn-

McMaster, I., J. R. Sampson, and J. E. King, “Computer Acquisi-
tion of Natural Language,” Int’l Jnl of Man–Machine Studies,


Miller, P. L., “An Adaptive Natural Language System that List-

Narasimhan, R., Modelling Language Behaviour, Springer–Ver-
lag, Berlin, 1981.

Powers, David M. W., “Experiments in Computer Learning of
Sydney NSW Australia, November 1984.

Powers, David M. W. and Christopher C. R. Turk, Machine
Learning of Natural Language, Springer, London/Berlin, De-
Australia, 1985.

Rayner, Manny, Asa Hugosson, and Goran Hagert, “Using a
logic grammar to learn a lexicon,” R88001, Swedish Institute

Rayner, Manny, “Applying Explanation-Based Learning to
Natural Language Processing,” R890144 and R890145, 1989.
Part 2 with Christer Samuelsson, also 1989.

Recker, Larry, “The computational study of language acquisi-

Recker, Larry, “Adaptive Individualized User Interfaces for
by the BDM Corp., McLean VA, for Applied Information
Technology Research Center, Columbus OH, 1988. Report 2

Rolandi, Walter G., “Language Acquisition by Machine: An Ope-
erant Investigation,” Masters Thesis, Dept. of Psychology,
Univ. 5th Carolina, 1988. (Supervisor: James B. Appel)

Salveter, Sharon C., “Inferring conceptual graphs,” Cognitive

Schank, Roger C. and Mallory Selfridge, “How to Learn/What to

Selfridge, Mallory, “A Computer Model of Child Language Ac-
quision,” 7th International Joint Conference on Artificial Intelli-

Sembagamoorthy, V., “PLAS, A Paradigmatic Language Sys-

Sembagamoorthy, V., Analog-based Acquisition of Utter-
ances relating to Temporal Aspects, 1981. Draft Submitted to
IJCAI-7.

Siklosy, Laurent, “A Language-Learning Heuristic Pro-

Siklosy, Laurent, “Natural Language Learning by Computer,”
in Representation and Meaning: Experiments with Informa-
tion Processing Systems, ed. H. A. Simon and Laurent Siklosy,


7.4. Significance
One of the first pieces of work which could claim to be MLNL is that of Yngve, cited in section 3.1. Contemporary is the work of Solomonoff, Spark-Jones and Lamb, with Machine Translation as the primary target, and statistical methods as the primary weapon. This sort of approach was, however, one of the main targets of the theoretical analyses of Gold and Chomsky (section 3) which showed that there were inherent problems with such simplistic approaches. Feigenbaum takes a more AI approach.

Following the example of Yngve, and parallel to the development of corresponding techniques in Connectionism, such techniques are now used in restricted (and normally lower) levels of the language hierarchy and attempt to capture some of the restrictions of human cognition. Such restrictions, as pointed out in section 3 and 4, are somewhat stronger than mere heuristics in that they actually define natural language. This leads to a change of perception, suggesting that similar methods are appropriate at corresponding levels of different modalities (suggesting extension to Ontology), and the different restrictions, teacher-critic characteristics and algorithms may be appropriate for learning at different levels of the hierarchies. Anderson and Powers have built preliminary implementations based on such ideas.

The work of Kelley, McMaster et al., and Harris are also historically significant pieces of work. Harris was one of the first to work with a deep semantics – "the parts of speech are the parts of the robot".

Another important contribution is the recognition of the place of errors – an important source of negative information, on the one hand, and a recognition that language is broader than textbook grammar, on the other. Langley, Powers and Lehman make use of errors rather than curtailing them, and Kelley introduced very early the idea of filtering out what did not fit into one's grammar and making use of the borderline, still comprehensible, cases for learning. This approach has been followed also by Reekers. The use of discrimination techniques is related, and essential, and has been pursued by these same researchers.

Some of the work aims to explore the use of a particular technique or approach. Berwick originally started off within Chomsky's TAG paradigm using the Marcus parser. Wirth and Rayner use particular specialized learning techniques in a language context as test bed, and Rayner and Samuelsson have, in particular, achieved impressive improvements in efficiency through the application of their approach. Salveiter and Selfridge worked in the context of Schank's Conceptual Dependency Graphs (section 2.2).

Applications have also called some projects into being. Rayner's we've mentioned. Zernik's, Wolf's, Wager's are also examples of this. Some work has had a very specialized focus, outside of the traditional preoccupation with grammar. Dreher and Kaye are concerned with Phonology, and Granger, Siskind and Sembagamoorthy (building on the approach of Narasimhan) with particular aspects of Semantics. There is a lot more room for work in what should not be peripheral areas.

7.5. Goals and Issues

Goal: The HAL of 2001, or Bridging the Communication Gap?

Issue: Most systems, and natural language learning experiments, are in danger of just translating from one representation to another. While this is appropriate for specific applications (database, machine translation, etc.), there is little merit in learning a one to one correspondence, or something close to it. Implementors need to make clear they are doing more than that.

Issue: Humans learn their language in parallel with their ontology! That is humans have to learn about their world too! A language learning system which cannot learn about its world is not adaptable, and has impaired language learning capability.

Issue: Most systems, and natural language learning experiments, start with simple examples of sentences (and/or meanings) and work up (if they're lucky) to complicated examples. Children learn primarily from full blown adult conversation. There is relatively little (machine readable) graded material. There is little advantage in constructing examples by rule. Learning is only possible of "what we almost already know". To use material which is beyond this "next grade" level, we need "filtering" – heuristic elimination of unprocessable input.
Issue: Some "field" systems provide mechanisms for accommodating to overly complex or new input, and optimizing to user variation and development. But the "too hard basket" is discarded. This, however, is precisely where learning systems should focus their effort, what is beyond the range of "acceptable" but nonetheless still "understandable". The excess baggage is never gratuitous!

Issue: What is the relationship between learning for recognition and learning for production? Children's generation capability seems to lag their understanding. Computers often reverse this trend!

Issue: Performance related learning is a factor in language learning, and a precursor to other aspects of language learning. But what role does it have and how can performance related developments in specialized domains incorporate into HALs.

Issue: Organization and consolidation have not been problems in some toy systems or specifically applied adaptive contexts. But in general, learning to associate similar things, classify and consolidate, can create problems in relation to memory. Programmers don't like to throw anything away. (It can involve implementational difficulties anyway.) People don't remember everything(?) And they certainly don't remember everything with the same ease or for the same time. Clutter can be a problem. The frame problem is really a special manifestation of this.

METRICS
1: Who provides the examples? (Teacher)

2: Who corrects the examples? (Critic)
3: Who evaluates the grammar? (Cheat)
4: How is meaning represented externally? (Examples)
5: How is meaning represented internally? (Knowledge)
6: What is the function of the system? (Interaction)
7: What aspects of grammar are learnt? (Phoneme to Book)
8: What aspects of semantics are learnt? (Noun to Article)
9: What aspects of ontology are learnt? (Robot or Database)

These are the metrics I have used in relation to the research listed. A comprehensive tabulation on the basis of such a list of metrics does not yet exist - I make an "impressionist" attempt above. One day...

8. Apologies

In conclusion, what more can I say? The field lies open! Who must I acknowledge? I must acknowledge all whose work I have cited here. And those whose work I have missed out on, misconstrued or undervalued? Please let me know! I'm sure you realize the impossibility of holding in one's head every detail of such a fast expanding and interdisciplinary area.

But I would appreciate it if those with relevant research and interests would contact me. And if you could characterize your own proposals and systems under the above metrics, so much the better.

A number of places will be available at the symposium for late registrations. Prospective participants can either contact the author on the Email address given above or approach the AAAI direct.