

**AI RESEARCH IN THE CONTEXT OF A
MULTIFUNCTIONAL KNOWLEDGE BASE:
THE BOTANY KNOWLEDGE BASE PROJECT¹**

Bruce W. Porter, Assistant Professor
James Lester, PhD student
Kenneth Murray, PhD student
Karen Pittman, Botanist and Knowledge Engineer
Arthur Souther, Post-doctoral researcher
Liane Acker, PhD student
Thomas Jones, Knowledge Engineer

SEPTEMBER 1988

AI88-88

¹ Support for this research was provided by the Army Research Office under grant ARO-DAAG29-84-K-0060 and the National Science Foundation under grant IRI-8620052.

Contents

Abstract.	2
1. Introduction.	3
2. Constructing the Botany Knowledge Base	4
2.1 The Focus of the Botany Knowledge Base	4
2.2 Lessons from the Knowledge-base Construction Project	5
3. Research in Machine Learning	7
3.1 The Task of Knowledge Integration	8
3.1.1 Recognition of New Information	8
3.1.2 Elaboration of New Information	9
3.1.3 Adaptation of Existing Knowledge and New Information	10
3.1.4 Psychological Basis of Learning as Knowledge Integration	11
3.2 An Application and Evaluation of Knowledge Integration.	11
3.2.1 An Example of Knowledge Integration	12
3.2.2 Recognition.	13
3.2.3 Elaboration.	15
3.2.4 Adaptation	16
3.2.5 Assessment of the Initial Results	17
Source of Power	17
Source of Complexity	17
Limitations	18
4. Research in Intelligent Tutoring	18
5. Summary	19
Acknowledgements	20
References	21
Appendix: An Overview of the Botany Knowledge Base	26

1. Introduction

We are engaged in research projects that explore the application of a large-scale, multifunctional knowledge base to significant AI problems. The premise of this work is the knowledge principle [LENA87, p 1173]:

“A system exhibits intelligent understanding and action at a high level of competence primarily because of the specific knowledge that it can bring to bear: the concepts, facts, representations, methods, models, metaphors, and heuristics about its domain of endeavor.”

While there is strong evidence for the knowledge principle in expert systems, very little research has explored the general applicability of a large-scale structured knowledge base to a broad range of cognitive tasks. Current expert systems draw on detailed knowledge of a narrow domain that is structured to support a single task (*e.g.*, identification [SHOR76]) and a single problem-solving method (*e.g.*, heuristic classification [CLAN85]). We believe that broader success in AI requires using multifunctional knowledge bases to further explore the knowledge principle. Our research in machine learning is an example of this exploration.

It is our premise that people learn by integrating new information into a wealth of existing knowledge. Although this conjecture seems obvious, few research projects have explored it. Most assume that the learner has little initial knowledge and that acquired knowledge improves problem-solving but not learning.¹ As discussed in Section 3, this assumption is contrary both to our intuitions about learning and to theories in developmental psychology.

To provide a “laboratory” in which to conduct our research, we are constructing a large-scale, structured knowledge base. We have chosen the domain of plant anatomy, physiology, and development because it is representative of nonformal domains yet is relatively self-contained. Like most domains, botany is concerned with objects (*i.e.*, plant anatomy) and processes that change objects (*i.e.*, plant physiology and development). Furthermore, botany incorporates both common sense and expert knowledge. For example, the common sense that living things require nutrients supports the expert knowledge that plant embryos consume endosperm. Since botany is not a formal domain, the knowledge is descriptive, heuristic, and incomplete. By contrast, knowledge in a formal domain is proscriptive, axiomatic, and comprehensive.

Although many researchers share our conviction of the knowledge principle, similar exploratory projects have been stalled by lack of resources. A primary frustration has been the “stone age” knowledge-base construction and editing tools that are available. We are constructing the Botany Knowledge Base using the excellent tools of CYC, which are under development by Doug Lenat’s group at

¹ See, for example, the general model of learning in [COHE82, pp 325–334].

MCC. Another frustration has been the *enormous* investments of time and effort required for construction. Each of the participants on our project has an individual research agenda that is predicated on exploring the knowledge principle; we are amortizing the expense of constructing the Botany Knowledge Base over a variety of projects.

This report summarizes the results of our first eighteen months of research activities. Section 2 describes our primary effort: construction of the Botany Knowledge Base. To summarize the (roughly) 4,000 frames, we present the top-level representations of objects and processes as well as the skeletal diagramming of the knowledge base. Sections 3 and 4 introduce the first two research projects that use the Botany Knowledge Base. Our first project studies knowledge integration, or “learning at the fringes” of existing knowledge. One anticipated outcome of this research is technology for semi-automatically extending a knowledge base [MURR88] (*i.e.*, a “next generation” Teiresias system [DAVI77]). Our second project concerns intelligent tutoring. In particular, we are exploring the dynamic generation of teaching plans and student-customized explanations of domain knowledge. We expect this research to improve intelligent tutoring by tailoring the presentation of knowledge to the particular needs of the student with more precision than is possible with current tutoring systems.

2. Constructing the Botany Knowledge Base

The primary activity during the first year of our research has been constructing the knowledge base. Section 2.1 describes the focus of the knowledge base and Section 2.2 discusses important lessons from the knowledge-base construction project.

2.1 The Focus of the Botany Knowledge Base

There are seven broad subareas in botany: anatomy, physiology, development, ecology, evolution, genetics, and taxonomy. We focus on the first three subareas:

- Anatomy is concerned with the component parts and subsystems of a plant. For example, important parts include the stamen of a flower, the tap root of the root system, and the chloroplasts of the shoot system. Crucial subsystems include the xylem network and the phloem network for distributing water, nutrients, and energy-containing compounds throughout the plant.
- Physiology is concerned with the roles played by anatomical and cellular components in processes important for life support, development, and reproduction of the plant. For example, important functions include photosynthesis, respiration, and hormonal regulation.
- Development is concerned with the changes affecting both the anatomy and physiology of a plant. Developmental changes must be viewed on both a small scale (*e.g.*, the growth and differentiation of a cell) and a large scale (*e.g.*, the development of a mature plant from an embryo).

Generally, we focus on botanical objects and the functional processes in which they participate.

The charts summarizing the Botany Knowledge Base are in the appendix. They contain the top-level representations of objects and processes, and the skeleton of the entire knowledge base.

2.2 Lessons from the Knowledge-base Construction Project

For the first three months of our project we attempted, with little success, to extract botany knowledge from textbooks. Our primary frustration was that textbook authors interleave multiple perspectives in a single description. Some of the perspective shifts were along the lines given above: the subarea of botany being emphasized and the developmental stage being described. Other perspective shifts were simplifications of domain knowledge. These “half-truths” are essential for introducing or summarizing complex concepts but must be reconciled with comprehensive perspectives.

More fundamentally, the implied relations among domain “facts” are often difficult to identify and incorporate. These relations include the consequences, implications, and support structure for each fact. The relations arise from the interactions between each new fact and the wealth of existing knowledge; they are crucial to integrating new information and to synthesizing a coherent overall understanding.

For example, consider the following fact about plant photosynthesis: *A plant uses light energy to perform photosynthesis and produce photosynthate.* Some of the relations between this new information and the pre-existing knowledge provide a support structure. These relations to existing knowledge (in bold print) might include:

- Photosynthesis is a **production process** with the plant as the **producer**, light as an **energy source**, and photosynthate as a **product**.
- Light energy is available in the **plant environment**.
- All **living things require food**, and plants produce their own food using light energy. This explains why a **plant dies if kept in the dark too long**.

Some of the relations elaborate the consequences of the new information. These relations might include:

- Since photosynthate is matter, raw materials are necessary to build it.
- The rate of photosynthesis will depend partly on the availability of light energy and raw materials.
- The plant must have some mechanism for assimilating (*i.e.*, acquiring, distributing, and utilizing) light energy and raw materials for photosynthesis.

- There may be waste products from photosynthesis, and the plant must have some mechanism for their disposal.

The search for relations between new information and existing knowledge can reveal inconsistencies. For example, the early development of a plant includes its embryo stage and emerging seedling stage. During these stages, the plant is without sunlight. Relating this new information to existing knowledge predicts that the young plant will die. Resolving this inconsistency requires additional knowledge of material-acquisition methods appropriate for young plants.

From our frustration with “direct knowledge transfer” we learned that thorough deliberation was required to represent each fragment of botanical knowledge. We identified inconsistencies and disambiguated meanings, often by identifying multiple perspectives on the same issue. This is essential to the flexible use of the knowledge base to support a range of cognitive tasks.

This flexibility is important to us. Most knowledge-based systems are inflexible because domain knowledge is included only if it is directly required for the problem-solving task; the support structure for the knowledge is extraneous. Crafting a knowledge base for a particular task, such as identification, sacrifices other tasks, such as tutoring and learning. This is similar to program compilation, which improves execution efficiency at the expense of modifiability and readability. As with program decompilation, introducing support structure in a knowledge base is very difficult.

The inflexibility of knowledge bases is apparent from current knowledge acquisition tools, which exploit the specificity of a “target” knowledge base in several ways. Some tools acquire knowledge by instantiating the model of a particular problem-solving method [ESH187] or task [MUSE87, BENN85]. Others acquire knowledge by debugging the knowledge base when problem-solving fails [DAVI77, BARE87]. Still others acquire knowledge by generalizing a set of examples to form classification rules [QUIN86, MICH80].

The Botany Knowledge Base is one of the first multifunctional knowledge bases. A multifunctional knowledge base is not committed to any particular problem-solving method or task. Rather, it encodes general knowledge of a domain to support diverse methods and tasks. For example, a multifunctional knowledge base for human physiology might be the substrate for expert programs that diagnose patients, tutor students, and organize reference materials. Automated tools for helping construct a multifunctional knowledge base cannot exploit knowledge-base specificity in the conventional ways.

The move to multifunctional knowledge bases is inevitable. The grid of potential knowledge bases has three dimensions: domain, task, and problem-solving method. Building knowledge-based systems for individual cells of this grid is both costly and shortsighted. Growing interest in multifunctional knowledge bases is evidenced by the first workshop on the topic [WORK88] and several other projects,

such as CYC [LENA87], Fermi [LARK88], and Kreme [ABRE86]. Further evidence is the use of general purpose knowledge representation languages and programming environments such as the CYC tools and language [SHEP88, LENA88], FrameKit [NIRE88], Opal [MUSE87], and Kee [INTE84]. These languages are not biased toward a particular problem-solving method. The next section describes our machine-learning research in knowledge acquisition for multifunctional knowledge bases.

3. Research in Machine Learning

Machine-learning research is motivated by two goals: constructing knowledge-based systems and modeling psychological theories of human learning [COHE82, CARB83, SIMO83]. However, because considerable effort has been devoted to the formalization of learning as concept formation (*e.g.*, [MITC82, QUIN83, MICH80, BARE87]), research results have demonstrated little progress toward these goals. We believe that progress toward the goals of machine learning requires the definition and exploration of tasks that better represent the general phenomenon of learning. We propose the task of knowledge integration – the incorporation of new information into existing knowledge.

Programs for concept formation offer negligible assistance with the construction of knowledge bases. These programs learn classification rules that summarize training examples to identify new examples. Research in knowledge-based systems is moving well beyond classification systems. As discussed above, current projects emphasize flexible support of multiple tasks.

Furthermore, concept formation, as formalized in machine learning, is of minor importance to theories of human learning. Learning tasks that assume little pre-existing knowledge *may* be important to theories of early development,¹ but theories of later learning emphasize the indispensable role of existing knowledge during learning. For example, psychological theories of concept formation and classification now emphasize the knowledge in which each concept is embedded [MURP85, p 289]:

“... current ideas, maxims, and theories concerning the structure of concepts ... are inadequate, in part, because they fail to represent intra- and inter-concept relations and more general world knowledge. We propose a different approach in which attention is focused on people’s theories about the world.”

A significant development in machine learning is recent work on explanation-based learning that demonstrates the essential role of domain knowledge in concept

¹ Researchers who believe concept formation is important to early development should demonstrate the adequacy of concept formation to learning “primitive concepts;” unfortunately, research primarily focuses on advanced concepts, such as disease categories [QUIN86, MICH87, KIBL87, BARE87], that no one contends are learned by people *via* concept formation. An exception to this focus is work on grammar learning (*e.g.*, [WHAR77]); however, the results so far give little confidence that inductive learning accounts for learning grammars of realistic size and complexity, and analysis of the task reveals its inherent limitations [GOLD78].

formation [MITC86, DEJo86]. Although we have gained considerable inspiration from this research, we note two implicit assumptions limiting its progress toward the two goals of machine learning. The first assumption is that an explanation is a deductive proof. This assumption is appropriate to a few artificial domains, but it ignores the complexity found in less formal domains and prevents knowledge-level learning [DIET87]. The second assumption is that the learner knows “just enough” to form a single explanation. This assumption avoids issues of weighing competing explanations, seeking unifying explanations, and controlling the search for an explanation. Ironically, the carefully crafted *lack of knowledge* is a crucial source of power for current explanation-based learning systems.

This section describes our initial efforts on research directed toward the goals of machine learning. A tenet of our research is that learning is knowledge intensive and involves establishing relationships between existing knowledge structures and new information. Section 3.1 introduces the task of knowledge integration, our steps toward a computational model, and the psychological basis for the learning task. Section 3.2 describes an application of the computational model for semi-automatically extending a knowledge base through discussion with a knowledge engineer.

3.1 The Task of Knowledge Integration

Knowledge integration is the incorporation of new information into existing knowledge. We believe that it is an active learning process involving three steps: recognition, elaboration, and adaptation. Recognition identifies known concepts that might relate to the new information. Elaboration establishes these relationships while filling-in missing details of the new information and identifying conflicts with existing knowledge. Adaptation modifies the learner’s knowledge to accommodate the elaborated information. Although it has received little attention in machine learning, knowledge integration is central to psychological models of learning.

3.1.1 Recognition of New Information

For new information to be learned, it must include references to concepts familiar to the learner [HAVI74]. These references provide indices into the learner’s existing knowledge to locate a place to record the new information. This is the process of *recognition*. For example, learning from the statement *The literacy rate of Iowa is 91%* requires knowledge of *Iowa*, *literacy rate*, and *91%*, and that *91%* is a sensible value for *literacy rate*, which itself is a sensible attribute of *Iowa*.

Sometimes the learner has multiple perspectives for each referenced concept. For example, perspectives for the concept *Iowa* might include: Iowa as a geological region, Iowa as an agricultural state, and Iowa as a political entity. When integrating new information about Iowa, the learner must determine the relevant perspective(s).

Therefore, recognition has two steps. When presented with new information, a learner must:

- 1) Identify existing knowledge structures corresponding to the referenced concepts.
- 2) Identify relevant perspectives for each referenced concept.

Identifying the relevant existing knowledge structures provides an interpretive context in which to integrate new information. The learner derives expectations from the knowledge structures. Applying these expectations to new information involves elaboration, which is discussed in the next section.

3.1.2 Elaboration of New Information

Elaboration is the embellishment of training; it results from the interaction of new information and existing beliefs that have been retrieved during recognition. Elaboration improves learning in three ways: it expands the information content of the training; it promotes consistency in the extended knowledge; and, by establishing more connections with prior knowledge, it enables new knowledge to be more accessible for subsequent use. This section discusses each improvement.

First, elaboration expands new information by relating it to existing knowledge. Gagne illustrates this with the following example [GAGN85, p 77]: A student is told *In vitro experiments show that Vitamin C increases the formation of white blood cells*. The student has prior knowledge that white blood cells destroy viruses and that Vitamin C is taken to fight colds, that are caused by viruses. Elaboration suggests that Vitamin C is capable of fighting colds *because* it stimulates creation of white blood cells, which subsequently kill cold-causing viruses. This conclusion was neither stated in the training nor previously known by the student. It arose from the student's effort to relate the new information to existing knowledge.

Second, elaboration identifies conflicts between new information and existing knowledge. For example, consider telling a learner that *President Nixon is a Quaker*.¹ This should trigger a greater response than simply recording Nixon's religious affiliation. From prior knowledge of Quaker pacifism, the learner predicts that Nixon is a dove. However, from prior knowledge of Nixon's defense policies, the learner predicts that Nixon is a hawk. Thus, elaboration reveals a conflict between new information and existing knowledge.

Third, elaboration makes new information more accessible for subsequent recall. For example, the student in Gagne's example might be reminded that Vitamin C increases white blood cell count by recalling the beliefs that Vitamin C fights colds and white blood cells destroy viruses that cause colds. The justification

¹ This example is from [TOUR87].

find reasons

STATEMENT: *This plant looks as if it has died.*

ELABORATION: *Of course it died; it wasn't producing food; all living things die without food.*

make predictions

STATEMENT: *Plants absorb water through their roots.*

ELABORATION: *The water is distributed through the xylem network.*

attribute details

STATEMENT: *The Hispaniola is a fine schooner.*

ELABORATION: *The Hispaniola probably has three masts, since it is a schooner.*

identify principles

STATEMENT: *Consumers compete to purchase the desired goods.*

ELABORATION: *As demand goes up, then price goes up too.*

Figure 1
Common Types of Elaboration

of Vitamin C fighting colds includes the belief that Vitamin C increases formation of white blood cells.

The elaboration of new information is potentially unbounded. A goal of our research is to identify frequently occurring classes of elaborations useful for the construction of knowledge-based systems, such as *attribution* (i.e., inheriting properties to a new concept from superordinate concepts) and *continuation* (i.e., predicting the outcome or temporal continuation of a process) [GAGN85]. A partial list of elaboration types is presented in Figure 1 (adapted from [GAGN85, WEIN78]).

3.1.3 Adaptation of Existing Knowledge and New Information

The final step in our learning model is to modify the learner's existing beliefs to accommodate the elaborated training. This involves generalizing the argument structures so that they apply to all sufficiently similar concepts. For example, generalizing the elaboration of the Nixon information might produce the belief that politicians' defense policies are more heavily influenced by political affiliation than by religion.

Modifications to the learner's knowledge are not restricted to generalizations of existing beliefs. Adding the generalized elaboration can introduce new relations, or shift the applicability of existing beliefs. Generalizing or specializing are simply two of many possible ways to shift the applicability of beliefs. In the Nixon example, the belief that religious preference determines defense policy was weakened in the case of politicians.

Contrast our approach to that of concept formation. Under concept formation, the statement *President Nixon is a Quaker* is a training instance in which President Nixon is a positive example of the concept Quaker. Explanation-based concept formation [MITC86, DEJO86] would try to confirm Nixon's religion, and then generalize the representation of Quaker with the explanation's weakest preconditions. Similarity-based concept formation would probably fail because the

example lacks relevant attributes, but it might conclude that presidents are Quakers, or that Nixons are Quakers. Neither learning method would reveal or resolve the assertion's irony.

3.1.4 Psychological Basis of Learning as Knowledge Integration

The view that learning involves relating new information to existing knowledge is fundamental to contemporary educational psychology. It is central to both Piaget's general learning theory of assimilation and accommodation [PIAG46], and to the Given-New Strategy for modeling learning from prose [HAVI74]. It is the foundation of the *learning strategy hypothesis*, a popular interpretation of numerous empirical studies, which claims [MAYE80, p 770]:

“activities aimed at making the learner actively integrate new information with existing knowledge affect the encoding, storage, and eventual use of new material on performance tests.”

This doctrine originated with the eighteenth-century philosopher and psychologist Herbart (1776-1841) [LANG93, DEGA95, RAND10, MILL79]. Herbart appears to be the first to explicitly advocate the process of relating new information to existing knowledge, which he called *apperception*, as the central activity of learning. Herbart notes that successful learning involves a double apperception: relating new observations to existing beliefs, and relating new explanations to existing laws and theories (*i.e.*, generalizing). These distinctions suggest the separate activities that we call recognition, elaboration, and adaptation.

Contemporary educational psychology also stresses the importance of relating new information to existing knowledge during learning [WEIN78]. Studies show that relations are constructed by elaborating new information [GAGN78]. Studies by Haviland and Clark indicate that readers elaborate to fill in missing details in text, and that this elaboration is performed as information is encountered, rather than only during recall, as in reconstructive memory [HAVI74]. Studies by Mayer [MAYE80] suggest two ways that elaboration enhances learning: by increasing a student's ability to recall information and by enabling the student to apply new information in solving problems. Reder [REDE79, REDE82B] has found that elaboration results in additional beliefs, either added to the learner's knowledge explicitly and directly accessible, or added implicitly and quickly computable.

3.2 An Application and Evaluation of Knowledge Integration

In this section, we discuss our application of knowledge integration to a knowledge acquisition tool that helps extend a knowledge base. This application is useful for two reasons. First, knowledge-base extension is a real problem confronting knowledge engineers, and its solution has inherent utility. Second, knowledge-base extension directly confronts the complexity of knowledge integration. The sources of this complexity are the numerous, subtle relationships between new information and existing knowledge that must be identified.

Our discussion centers on an example of knowledge integration that is representative of the complex task of extending a knowledge base. A prototype implementation of our knowledge integration theory, called ProtoKI, has been completed and has been successfully tested with this learning scenario. This example has been simplified for presentation. For example, ProtoKI does not generate and parse natural language; this discourse has been converted from a language of frames, slots, and values. We provide a complete description of the prototype and the implementation of this learning scenario in [MURR88].

In our discussion of the ProtoKI program, we first present the learning example as a dialogue between ProtoKI and a knowledge engineer. Then, in Sections 3.2.1 through 3.2.3, we discuss this example in terms of recognition, elaboration, and adaptation.

3.2.1 *An Example of Knowledge Integration*¹

This example involves extending the Botany Knowledge Base with new information about plant seeds. The knowledge base already has information that plant seeds contain nutritive tissue called endosperm, which is analogous to an egg's yolk. The plant embryo consumes the endosperm during its development inside the seed. A knowledge engineer wishes to extend the knowledge base with a representation of *nonendospermic seeds*, a type of seed that contains no endosperm. The task of ProtoKI is to interact with the knowledge engineer to integrate this new information.

The knowledge engineer presents new information to ProtoKI:

Knowledge Engineer: "There is a class of seeds that have no endosperm."

ProtoKI retrieves knowledge structures relevant to this new information. Using inference rules defined for the retrieved knowledge structures, ProtoKI forms the expectation that the embryo in a nonendospermic seed will die of starvation. It then attempts to confirm this expectation with the knowledge engineer:

ProtoKI: "These seeds do not give rise to healthy seedlings."

Knowledge Engineer: "Explain."

ProtoKI: "The embryo starves since the endosperm provides nutrients, and nutrients are essential for health and survival."

Knowledge Engineer: "Yes, nutrients are essential; however, the embryo survives."

¹ This section forms the basis of a paper appearing in the Proceedings of the Knowledge Acquisition for Knowledge-based Systems Workshop, Banff, Canada, November 6-11, 1988.

ProtoKI attempts to resolve the inconsistency between its expectations and the assertion that embryos in nonendospermic seeds survive. One possible explanation is that these embryos acquire nutrients from some other source. ProtoKI searches for alternate ways an embryo-stage plant might acquire nutrients. A focused discussion with the knowledge engineer ensues that resolves the inconsistency and extends the knowledge base.

ProtoKI: “Does the embryo acquire nutrients from photosynthesis?”

Knowledge Engineer: “Yes, the shoots of some plant embryos emerge from the seed during germination and start producing photosynthate.”

As this example illustrates, ProtoKI integrates new information by determining its consequences. When inconsistencies are encountered, ProtoKI suggests possible explanations to resolve them. This guides the acquisition of further knowledge.

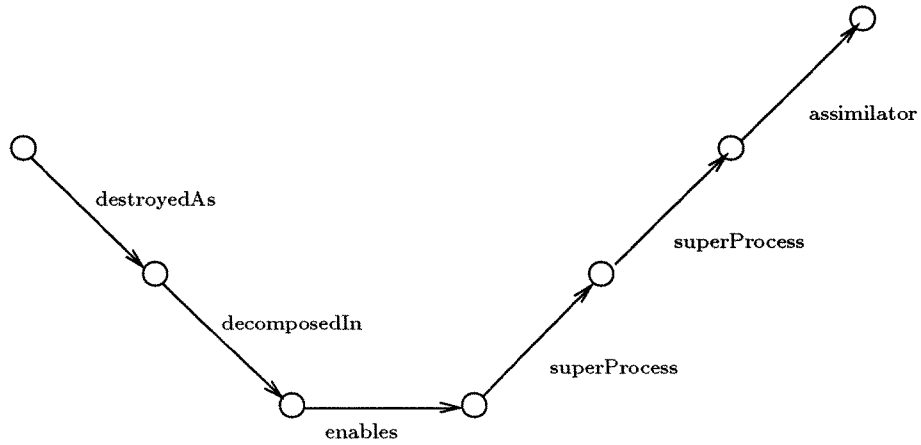
The computational issues that arise during knowledge integration include identifying knowledge relevant to new information, relating new information to the relevant knowledge, and adapting the knowledge base to accommodate the new information. The following three sections describe in greater detail how ProtoKI performs these activities.

3.2.2 Recognition

ProtoKI begins knowledge integration by identifying relevant knowledge structures. In the previous example about seeds with no endosperm, ProtoKI must determine which among the thousands of frames in the Botany Knowledge Base may be affected in some interesting way.

To focus the search for knowledge relevant to new information, the representation of each object in the Botany Knowledge Base is structured with **perspectives**. Each perspective is a partitioning of the knowledge base that describes an object in a particular role. For example, one perspective of endosperm is *Plant Food Source*, as shown in Figure 2. Other perspectives include: endosperm as a *Product Of Reproduction*, endosperm as a *Contained Object*, and endosperm as a *Taxon Defining Part*. ProtoKI collects the perspectives for objects referenced by new information and prompts the knowledge engineer to select which are appropriate.

A perspective is a semantic-net template that can be instantiated for hypothetical objects. The instantiation of *Plant Food Source* for an endosperm is presented in Figure 3. Collectively, these instantiated frames comprise a **context** representing an endosperm in its role as a plant food source; this context is used to simulate the effects of the new information about endosperm.



This semantic-net template defines the context relevant to an object in its role as a plant food source: A plant food source must have a stage when it is destroyed and decomposed into nutrients. This decomposition enables the nutrients to be assimilated by the plant. The nested subprocesses of assimilation are provision and acquisition.

Figure 2
The Perspective *Plant Food Source*

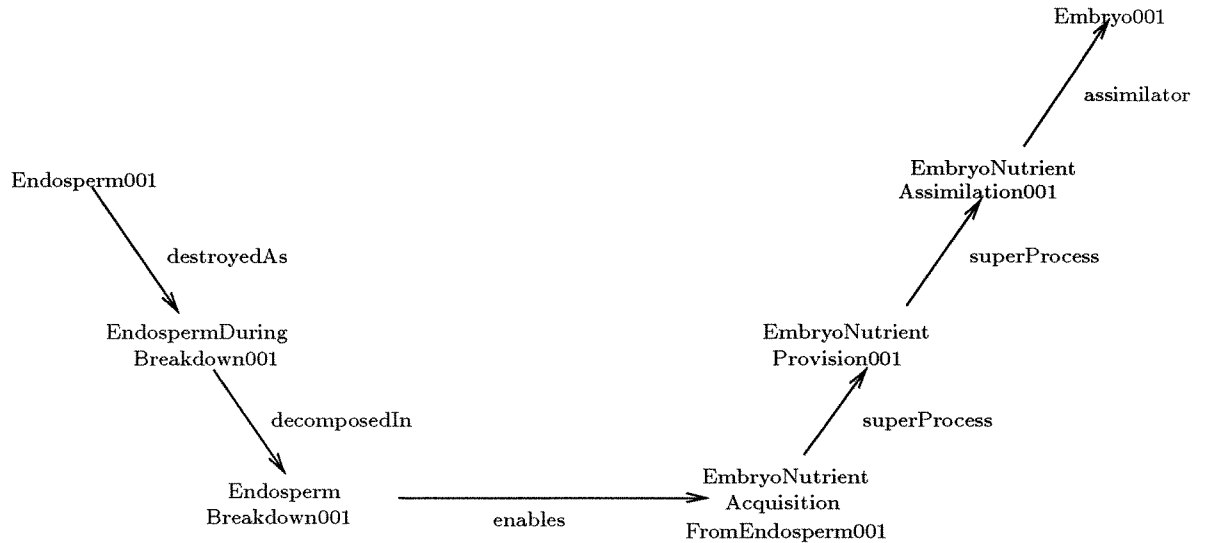
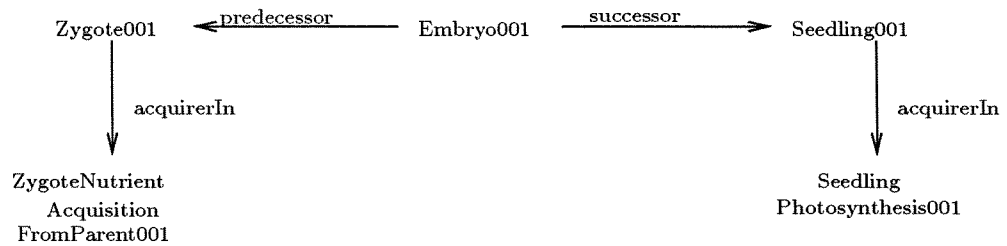


Figure 3
The Context Created by Instantiating *Plant Food Source*

-
1. When an entity is disabled, all of its developmental stages are disabled.
 2. When an entity is disabled, all the processes involving the entity are disabled.
 3. When a process is disabled, all the processes that its completion enables are disabled.
 4. When the known methods of acquiring some essential resource are disabled, the rate of provisioning the resource is inadequate for survival.
 5. When the assimilation rate for some resource is unknown, it is the same as the provision rate.
 6. When nutrient assimilation is inadequate for survival, the assimilator is dying.
 7. When a living entity is dying, the succeeding developmental stage is disabled.
-

Figure 4
Heuristic Rules Relevant to Endosperm as a Plant Food Source



This is the extension to the context of Figure 2, showing the developmental predecessor and successor of *Embryo 001* and their methods of nutrient acquisition.

Figure 5
The Context Extension

3.2.3 Elaboration

During recognition, ProtoKI has isolated the context in the knowledge base most relevant to the new information. Next, during elaboration, ProtoKI determines how the new information interacts with existing knowledge within this context. Elaboration involves applying inference rules to propagate the effects of the new information throughout the context.

In the endosperm example, elaboration begins when ProtoKI asserts that the endosperm is absent from the context by assigning value *False* to the slot *enabled?* of *Endosperm 001*. This assignment triggers inference rules that determine the consequences of seeds lacking endosperm. For example, without the endosperm, the embryo cannot get enough nutrients to survive. The inference rules applicable to this example are listed in Figure 4.

Through elaboration, ProtoKI concludes that the plant embryo is dying from lack of nutrients. This triggers the instantiation of a second semantic-net template defined for plants that are starving and in danger of dying. The context is expanded to include the plant's developmental stages immediately before and after its embryo

stage and how nutrients are acquired during each of these developmental stages. Through continued elaboration, ProtoKI concludes that the plant’s seedling stage is not reached because the plant dies during its embryo stage. The context extension is presented in Figure 5.

An important function of elaboration is identifying inconsistencies. These occur when expectations of the knowledge base are violated by new information or when two rules reach conflicting conclusions. Resolving inconsistencies involves correcting the new information to comply with current expectations or adapting the existing knowledge structures to accommodate the new information.

3.2.4 Adaptation

Elaboration can reveal inconsistencies in the knowledge base; the adaptation step resolves them. An inconsistency can be introduced by inference rules used during elaboration or by facts the knowledge engineer asserts. In the endosperm example, an inconsistency is detected when the knowledge engineer asserts that the embryos of nonendospermic seeds survive, correcting the prediction that these embryos starve.

Resolving inconsistencies requires correcting explanations that support failed expectations and constructing alternative explanations to account for new information. When the knowledge engineer refutes the prediction that embryos of nonendospermic seeds starve, ProtoKI inspects the explanation for this prediction to determine its weakest premise. Rule 4 (from Figure 4) relies on a closed-world assumption and is considered a relatively weak inference. Therefore, ProtoKI retracts its conclusion and assumes *Embryo Nutrient Provision 001* is adequate for the embryo’s survival. This change propagates through the explanation, retracting the belief that the embryo starves.

The original inconsistency has been resolved by assuming adequate nutrient provision by the embryos of nonendospermic seeds. However, no alternative method is known for plant embryos to acquire nutrients. ProtoKI seeks to construct an explanation for the assumed nutrient acquisition using the following inference:
 IF a resource provision is adequate for survival, but no acquisition method is known
 THEN assume the acquisition method of the developmental successor is employed.

This rule suggests the embryos of nonendospermic seeds acquire nutrients by photosynthesis, as is done by seedlings. However, this hypothesis introduces new constraints on the embryos of nonendospermic seeds. For example, to acquire nutrients by photosynthesis, the embryo must be a photosynthetic plant. Therefore, to apply this inference, ProtoKI asserts that *Embryo 001* is an instance of *Photosynthetic Plant*. As a photosynthetic plant, the embryo inherits the following features: its color is green, it contacts sunlight, and its composition includes chlorophyll. In short, the plausibility of explaining the survival of nonendospermic embryos by assuming they engage in photosynthesis is contingent on their contacting sunlight

and containing chlorophyll. Confirming these assumptions leads to the acquisition of further knowledge from the knowledge engineer.

3.2.5 *Assessment of the Initial Results*

The preceding example is representative of the complex knowledge-base extensions that a knowledge integration tool should handle. Implementing this scenario has provided several insights into the task of knowledge integration and our approach to automating it. In this section, we highlight the lessons derived from this exercise.

Source of Power

KI uses the domain expertise of the knowledge base to determine the consequences of new information. By contrast, other approaches to critiquing knowledge-base extensions underutilize the existing knowledge. These include:

- Teiresias [DAVI77] critiques new rules with *rule models*. Rule models record correlations between antecedent terms and consequent terms of the existing rule base. If a new rule violates a rule model, Teiresias suggests how the rule can be modified to conform to the model.
- Knac [LEFK87] anticipates knowledge-base revisions with discourse cues and a heuristic theory of knowledge acquisition. Knac is capable of predicting which structures to revise when new information is ambiguous.
- Odysseus [WILK88] validates new rules with a *confirmation theory*. This involves techniques to assess the quality of a new rule for subsequent problem solving and to ensure its consistency against a set of test cases.

Each of these methods demonstrates the use of external knowledge or meta-knowledge to evaluate new information. However, they cannot detect subtle inconsistencies introduced by knowledge-base revisions because they ignore the consequences of new information on existing knowledge.

Source of Complexity

The first step of knowledge integration requires identifying existing knowledge that might be affected by new information. This task is quite difficult because of four complications. First, the inconsistencies introduced by knowledge-base revisions are often implicit. Revealing them requires determining the consequences of the new information using the system's inference rules. However, computing the deductive closure of a knowledge base is intractable. Second, the network paths relating two frames may be quite long. Since these may lead to the "deep" consequences of the new information, the paths must be pursued. Third, the knowledge-base frames are densely interconnected. This precludes unfocused search methods, such as spreading activation. Finally, since no fixed performance task is assumed, we cannot simply consider the consequences of new information for a fixed set of queries.

Limitations

Our implementation of a tool for knowledge integration ignores the issue of selecting perspectives. For the next phase of development, we are designing an agenda-based architecture to address this issue. Tasks on the agenda select perspectives and instantiate them. A perspective is selected for possible instantiation when its degree of overlap with the current context exceeds an activation threshold. Selected perspectives are ordered for instantiation by a heuristic measure of “interestingness.” This is a function of the conflict level of inconsistencies contained in those portions of the context that each perspective overlaps. This control mechanism permits KI to conduct a best-first search through the deductive closure of the extended knowledge base, either autonomously or under the user’s guidance.

Because knowledge integration assumes substantial domain knowledge, tools for knowledge integration are limited. In particular, they are inappropriate during the initial stages of knowledge-base development when the encoded domain expertise is sparse.

4. Research in Intelligent Tutoring

Our second major project exploiting the Botany Knowledge Base concerns intelligent tutoring. In particular, we are interested in knowledge representations and processes for adapting teaching plans and explanations to the individual needs of a student. The rigidity of past tutoring programs is due largely to the lack of adaptable teaching plans and explanations. These depend on a rich knowledge of the domain, the student’s understanding, and the available teaching mechanisms. This section describes our proposed research and expected contributions, although the research is just underway.

The goal of tutoring is improving the student’s understanding of a particular domain. To achieve this goal the tutor must have a plan for presenting knowledge in an organized fashion. An important part of this presentation is a set of standard explanations that provide cohesion to the knowledge. Ideally, the tutor need never change the teaching plan or the form of these standard explanations during the tutorial session. However, in most sessions, the teacher cannot predict in advance the student’s understanding at each stage of the interaction. Therefore, good tutors can dynamically revise their teaching plans and generate novel explanations in response to changing perceptions of the student’s knowledge.

Clearly, an Intelligent Tutoring System (ITS) would benefit from the ability to adjust to the student’s current needs through the use of dynamic pedagogical planning and explanation generation. However, most ITSs either ignore teaching plans and explanations entirely, or they employ ones that are preformulated. This means that teaching plans cannot be dynamically modified to cope with unanticipated student needs. Similarly, preformulated explanations lack sensitivity to changing student needs and fail to take advantage of the inferential power of the domain representation.

The goal of the proposed research is to construct an ITS that provides two fundamental enhancements to the current state of the art: dynamic pedagogical planning and student-customized explanation generation. Because the enhancements depend critically on a sophisticated student model, a secondary goal is developing a student modeling facility that provides a richer representation than those in current ITSs.

We intend to develop a dynamic planning facility that constructs plans for teaching episodes. During the course of a tutoring session, some plans may fail to achieve the intended goals, while others may be rendered unnecessary, either because of certain changes in the student's understanding or because of changes in higher-level goals. In these cases, the system will be able to dynamically revise plans to respond to the idiosyncratic flow of each session. Plan revision will be based on the current contents of the student model, curricular information embedded in the domain model, and a library of pedagogical strategies.

We intend to develop a dynamic explanation facility for constructing explanations that should be more effective than those in current systems for two reasons. First, the explanations will be generated directly from the representation of the domain knowledge. This enables the system to provide a variety of explanations expressing different shades of meaning. Second, they will be more customized to the student because the explanation generator will have access to a rich student model. This allows the presentation of explanations that are more easily understood by the student. Since the system can generate a multitude of explanations for a given phenomenon, and it can use a more precise understanding of the student's current knowledge state to generate the explanations, the resulting explanations should address the student's needs more specifically.

In combination, a dynamic planner and explanation generator should improve student learning at two levels. At the global level, the dynamic planner can react to changing student needs and interests, such as resolving a mismatch in perspectives between the student and the tutoring system. In contrast, the explanation generator operates at the local level, e.g., correcting a misunderstanding about a particular concept. The net effect of this multilevel interaction should be a significantly more successful and efficient presentation to the student.

5. Summary

Our research explores the application of a large-scale, multifunctional knowledge base to a variety of AI tasks. During the past year we have constructed a knowledge base that, we believe, will flexibly support a range of cognitive modeling experiments.

Our knowledge base is in the domain of botany. Our focus is the anatomy, physiology and development of plants, with emphasis on structured objects and the processes causing their creation and change. We are constructing our knowledge base using the excellent tools of CYC which are under development by Doug Lenat

at MCC. The enormous investment of time and effort devoted to construction is being amortized over two research projects now, and more that are planned for the future.

Currently, we are studying machine learning and tutoring. The goal of our learning research is to develop and evaluate a computational model for a new learning task – knowledge integration. This task, which might be termed “learning at the fringes of a knowledge base,” involves the incorporation of new information into existing knowledge. The goals of machine learning – knowledge acquisition and cognitive modeling – emphasize the importance of this learning task and suggest that research on concept formation, the predominant concern of machine learning research, is of limited applicability. We are applying our computational model of this task to the problem of automated knowledge acquisition, in particular the extension of a knowledge base through interaction with a knowledge engineer.

The goal of our intelligent tutoring research is to develop knowledge representations and processes for dynamically adapting teaching plans and explanations to the individual needs of a student. Most current intelligent tutoring systems either ignore teaching plans and explanations entirely, or they employ ones that are pre-formulated. As a result, teaching scenarios cannot be dynamically modified to cope with unanticipated student needs. The ability to dynamically modify a tutoring session is critically dependent on rich domain knowledge.

Our work is predicated on the knowledge principle that emphasizes the crucial role in intelligence of extensive, task-independent knowledge. This research cannot be done “in the small,” and we are committed to its long-term exploration. Although we are in the early stages of research, the significance of the knowledge principle has already become evident.

Acknowledgements

We are indebted to Doug Lenat for inspiring our research on the knowledge principle and creating the CYC knowledge representation language and knowledge-base editing tools. We appreciate the help of all the members of the CYC group, especially R.V. Guha, David Wallace, John Huffman, Mary Shepard, and Claudia Porter who helped develop and stabilize the CYC software. We appreciate the community of researchers in machine learning at the UT-Austin, AI Lab, in particular Rob Holte, Ray Mooney, Ray Bareiss, Brad Blumenthal, Karl Branting, Rich Mallory and Agnar Aamodt. Finally, we appreciate MCC’s donation of our computer equipment and the open research environment that permits our work to be freely discussed and distributed.

REFERENCES

- [ABRE86] Abrett, G. and Burstein, M.H. The KREME Knowledge Editing Environment. *Proceedings of the Knowledge Acquisition for Knowledge-based Systems Workshop* (1986).
- [BARE87] Bareiss, E.R., Porter, B.W. and Wier, C.C. Protos: An Exemplar-Based Learning Apprentice. *Proceedings of the Fourth International Workshop on Machine Learning* (1987), 12-23.
- [BENN85] Bennett, J.S. ROGET: A Knowledge-Based System for Acquiring the Conceptual Structure of a Diagnostic Expert System. *Automated Reasoning 1:1* (1985), 49-74.
- [CARB83] Carbonell, J.G., Michalski, R.S., and Mitchell, T.M. An Overview of Machine Learning. Appearing in *Machine Learning: An Artificial Intelligence Approach*, Tioga Publishing, 1983, pp. 3-24.
- [CLAN85] Clancey, W.J. Heuristic Classification. *Artificial Intelligence* 27, 289-350.
- [COHE82] Cohen, P.R. and Feigenbaum, E.A. *The Handbook of Artificial Intelligence*, William Kaufmann Publ., 1982.
- [DAVI77] Davis, R. Interactive Transfer of Expertise: Acquisition of New Inference Rules. *Proceedings of the International Joint Conference on Artificial Intelligence* (1977), 321-328.
- [DEGA95] De Garmo, C. *Herbart (and the Herbartians)*, New York: Charles Scribner's Sons, 1895.
- [DEJo86] DeJong, G. and Mooney, R. Explanation-Based Learning: An Alternate View. *Machine Learning* 1, 2 (1986), 145-176.
- [DIET87] Dietterich, T.G. A Knowledge Level Analysis of Learning Programs. (Technical Report 87-30-4 from the Computer Science Department, Oregon State University).
- [ESHL87] Eshelman, L., Ehret, D., McDermott, J. and Tan, M. MOLE: A Tenacious Knowledge Acquisition Tool. *International Journal of Man-Machine Studies* 26, 1 (1987), 41-54.

- [GAGN78] Gagne, E.D. Long-term Retention of Information Following Learning from Prose. *Review of Educational Research* 48, 4 (1978), 629–665.
- [GAGN85] _____. *The Cognitive Psychology of School Learning*, Boston: Little, Brown and Company, 1985.
- [GOLD78] Gold, E.M. Complexity of Automaton Identification from Given Data. *Information and Control*, Volume 37 (1978), 302–320.
- [HAVI74] Haviland, S.E. and Clark, H.H. What’s New? Acquiring New Information as a Process in Comprehension. *Journal of Verbal Learning and Verbal Behavior* 13 (1974), 512–521.
- [INTE84] KEE Software Development System. A commercial product of Intellicorp, Mountain View, California.
- [KIBL87] Kibler, D. and Aha, D.W. Learning Representative Exemplars of Concepts: An Initial Case Study. *Proceedings of the Fourth International Workshop on Machine Learning* (1987), 24–30.
- [LANG93] Lange, K. *Apperception: A Monograph on Psychology and Pedagogy*, De Garmo, C. (trans. ed.), D. C. Heath and Company, 1893.
- [LARK88] Larkin, J.H., Reif, F., Carbonell, J. and Gugliotta, A. FERMI: A Flexible Expert Reasoner with Multi-Domain Inferencing. *Cognitive Science* 12 (1988), 101–138.
- [LEFK87] Lefkowitz, L.S. and Lesser, V.R. Knowledge Acquisition as Knowledge Assimilation. *Proceedings of the 2nd Knowledge Acquisition for Knowledge-Based Systems Workshop* (1987), 14.0–14.13.
- [LENA87] Lenat, D.B. and Feigenbaum, E. On the Thresholds of Knowledge. *Proceedings of the International Joint Conference on Artificial Intelligence* (1987), 1173–1182.
- [LENA88] Lenat, D.B. and Guha, R.V. The World According to CYC. Technical Report ACA-AI-300-88, MCC, Austin, Texas.
- [MAYE80] Mayer, R.E. Elaboration Techniques That Increase the Meaningfulness of Technical Text: An Experimental Test of the Learning Strategy Hypothesis. *Journal of Educational Psychology* 72, 6 (1980), 770–784.

- [MICH80] Michalski, R.S. and Chilausky, R.L. Learning by being Told and Learning from Examples: an Experimental Comparison of the Two Methods of Knowledge Acquisition in the Context of Developing an Expert System for Soybean Disease Diagnosis. *Policy Analysis and Information Systems*, Vol. 4, No. 2 (June, 1980), 125–160. (Special issue on knowledge acquisition and induction).
- [MICH87] Michalski, R.S. How to Learn Imprecise Concepts: A Method for Employing a Two-Tiered Knowledge Representation in Learning. *Proceedings of the Fourth International Workshop on Machine Learning* (1987), 50–58.
- [MILL79] Miller, G.A. Images and Models, Similes and Metaphores. Appearing in *Metaphore and Thought*, Cambridge University Press, 1979.
- [MITC82] Mitchell, T.M. Generalization as Search. *Artificial Intelligence*, Volume 18 (1982), 203–226.
- [MITC86] Mitchell, T.M., Keller, R.M., and Kedar-Cabelli, S.T. Explanation-Based Generalization: A Unifying View. *Machine Learning* 1, 1 (1986).
- [MURP85] Murphy, G.L. and Medin, D.L. The Role of Theories in Conceptual Coherence. *Psychological Review* 92, 3 (1985), 289–316.
- [MURR88] Murray, K. KI: An Experiment in Automating Knowledge Integration. PhD Dissertation Proposal, Department of Computer Science, University of Texas at Austin (forthcoming Technical Report).
- [MUSE87] Musen, M.A., Fagan, L.M, Combs, D.M. and Shortliffe, E.H. Use of a Domain Model to Drive an Interactive Knowledge-editing Tool. *International Journal of Man-Machine Studies* 26 (1987), 105–121.
- [NIRE88] Nirenburg, S., Monarch, I., Kaufmann, T., Nirenburg, I. and Carbonell, J. Acquisition of Very Large Knowledge Bases: Methodology, Tools and Applications. Forthcoming Technical Report from Carnegie-Mellon University.
- [PIAG46] Piaget, J. Piaget’s Theory. Appearing in *Carmichael’s Manual of Child Psychology*, Mussen, Paul (ed.), John Wiley and Sons, Inc., 1946, pp. 703–732.

- [QUIN83] Quinlan, J.R. Learning Efficient Classification Procedures and their Application to Chess End Games. Appearing in *Machine Learning: An Artificial Intelligence Approach*, Michalski, R.S., Carbonell, J.G., and Mitchell, T.M. (eds.), Tioga Publishing, 1983, pp. 463–482.
- [QUIN86] ———. Induction of Decision Trees. *Machine Learning* 1,1 (1986), 81–106.
- [RAND10] Randels, G.B. *The Doctrines of Herbart in the United States*, PhD Dissertation, Department of Psychology, University of Pennsylvania, 1910.
- [REDE79] Reder, L.M. The Role of Elaborations in Memory for Prose. *Cognitive Psychology* 11 (1979), 221–234.
- [REDE82B] Reder, L.M. Plausibility Judgements versus Fact Retrieval: Alternative Strategies for Sentence Verification. *Psychological Review* 89 (1982), 250–280.
- [SHEP88] Shepherd, M. Tools for Adding Knowledge to the CYC Large-Scale Knowledge Base. MCC Technical Report No. ACA-AI-068-88.
- [SHOR76] Shortliffe, E.H. *MYCIN: Computer-based Medical Consultations*, New York: Elsevier, 1976. (Based on PhD Dissertation, Computer Science Department, Stanford University, 1974.)
- [SIMO83] Simon, H.A. Why Study Machine Learning?. Appearing in *Machine Learning: An Artificial Intelligence Approach*, Tioga Publishing, 1983, pp. 25–38.
- [TOUR87] Touretzky, D.S., Horty, J.F. and Thomason, R.H. A Clash of Intuitions: The Current State of Nonmonotonic Multiple Inheritance Systems. *Proceedings of the Tenth International Joint Conference on Artificial Intelligence* (1987), 476–482.
- [WEIN78] Weinstein, C.E. Elaboration Skills as a Learning Strategy. Appearing in *Learning Strategies*, O’Neil, H.F. Jr. (ed.), Academic Press, 1978, pp. 157–209.
- [WHAR77] Wharton R.M. Grammar Enumeration and Inference. *Information and Control* 33 (1977), 253–272.

- [WILK88] Wilkins, D.C. Knowledge Base Refinement Using Apprenticeship Learning Techniques. *Proceedings of the National Conference on Artificial Intelligence* (1988), 646–651.
- [WORK88] AAAI Workshop on Very Large Multifunctional Knowledge Bases. (Held at the National Conference on Artificial Intelligence, St. Paul, MN, August 22, 1988. Organizing Committee: Carbonell, J. Rich, E., Thomason, R., Nirenburg, S. and Monarch, I.).