LANGUAGE, POLITICS, AND METHOD:
GROUNDING A COMPUTATIONAL HERMENEUTIC

by

GAVAN DUFFY

B.A., Communication
University of Houston
(1979)

M.S., Political Science
Massachusetts Institute of Technology
(1984)

Submitted to the Department of Political Science
in Partial Fulfillment of
the Requirements for the Degree of
Doctor of Philosophy in Political Science

at the

Massachusetts Institute of Technology

August, 1987

(c) Gavan Duffy 1987

The author hereby grants to MIT permission to reproduce and to
distribute copies of this thesis document in whole or in part.

Signature of Author ____________________________

Department of Political Science
August 12, 1987

Certified by _______________________________

Hayward R. Alker, Jr.
Professor, Political Science
Thesis Supervisor

Accepted by ____________________________

Harvey M. Sapolsky
Chair
Graduate Program Committee

ARCHIVES
MASSACHUSETTS INSTITUTE
OF TECHNOLOGY
MAR 14 1988
LANGUAGE, POLITICS, AND METHOD:

GROUNDING A COMPUTATIONAL HERMENEUTIC

by

GAVAN DUFFY

Submitted to the Department of Political Science
on June 10, 1987 in partial fulfillment of the
requirements for the Degree of Doctor of Philosophy in
Political Science

ABSTRACT

This dissertation presents the results of part of an ongoing,
collaborative research project aimed at the implementation of a
computational environment for the analysis of political texts.
Additionally, it presents theoretical motivations for many of the
implementational choices made in the development of the system. Three
system components are described. One, a syntactic parser, is installed
in the system. Two others, a lexicon shell and a temporal indexation
mechanism, have been implemented in prototype form.

The research was motivated in part by a desire to overcome the
validity limitations of existing quantitative approaches to political
content analysis. The approach adopted is "computational
hermeneutics", which is distinguished here from other recent usages.
The theoretical motivation for the syntactic model is discussed by
means of a review of alternative theories of language acquisition. It
is argued that the acquisition of knowledge, including ones knowledge
of language, involves the progressive elaboration of conceptual
structures that encode the necessary and sufficient conditions, plus
\textit{o} \textit{e} \textit{t} \textit{e} \textit{r} \textit{e} \textit{i} \textit{s} \textit{ p} \textit{a} \textit{r} \textit{i} \textit{b} \textit{i} \textit{s}\ exceptions, for category membership. These structures
are shown to support the formation of hypotheses, including novel
hypotheses not directly abducible from existing theoretical constructs.

It is argued that the architecture of categories and concepts varies
across interpreters, reflecting variations in the life-histories of
each. Capturing these variations is important for capturing the
cognitive incommensurabilities across interpreters whose beliefs and
political interests conflict. Conceptual Dependency theory in
artificial intelligence and the theory of the lexicon in linguistics
are criticized for presuming conceptual invariability across
interpreters and discourse contexts.

\textbf{Thesis Supervisor:} Dr. Hayward R. Alker, Jr.
\textbf{Title:} Professor of Political Science
Acknowledgements

Ordinarily, one acknowledges the forebearance of one's family last, but Lily H. M. Ling must be acknowledged first and foremost for her tireless graciousness in listening to endless midnight flaming sessions about various details of Relatus. As well, she must be acknowledged for having the good sense to conduct field research in China while this manuscript was being prepared.

John C. Mallery, who has collaborated with me for about five years, deserves a great deal of credit for any useful ideas embedded within this text. From the dialectic of our close interaction a research program has emerged that is likely to take us both much further. Although he does not know it, Lloyd S. Etheredge deserves our thanks for convincing us to tackle the hard problems.

Helpful critical comments on various chapters of this manuscript were provided by Erik A. Devereux, Robert J. P. Ingrina, Lily H. M. Ling, John C. Mallery, Benjamin I. Page, Jim Sidanius, and Robert P. Weber. Rick Lathrop contributed to early thoughts that led to the temporal indexation mechanism. Conversations with Roger Hurwitz, David Kirsh, and Lee Mintz helped keep me from entering a variety of conceptual cul-de-sacs over the past few years.

The intellectual and, occasionally, financial support of the members of my committee, Hayward Alker, Tom Ferguson, and Marvin Minsky, made this work possible. I must also acknowledge the support of my original advisor, Ithiel de Sola Pool, who encouraged the direction of this research in its early days. I wish that Professor Pool had lived to see this research at its current stage.
Certain institutions also contributed to this work. The MIT Artificial Intelligence Laboratory supported my work for two academic years with a research assistantship and seemingly infinite computational resources. I thank Marvin Minsky, Patrick Winston, and Carl Hewitt for this. The University of Texas Artificial Intelligence Laboratory is also owed a debt of gratitude for allowing me to use their facilities when needed. I thank Gordon Novak, Martin Purvis, and Robert Simmons of the UT AI Lab. Symbolics, Inc., supported the development of the GEL Shell (Chapter 7) at their Cambridge Research Center during the Summer of 1986. Additionally, Symbolics has loaned me one of their 3600-series computers on which I can continue this research. I thank John L. Kulp and Howie E. Shrobe of Symbolics for this support. The temporal indexation mechanism was conceived during the Summer of 1985 at the CAD Research unit of Gould Electronics' Semiconductor Division. I thank Bob Kirk of Gould.

Corinne Hebda of the University of Texas Department of Government performed the Herculean task of organizing the final versions of the manuscript. She is to be thanked for putting up with endless minor revisions and for handling my occasionally frantic neuroses. Others who assisted when the need arose were Jack Byrd, Diana Jones, Marti Martin and Brian Roberts. Sue Burke and Mike Munger helped keep me sane. Thanks to you all.

Finally, I must express my deepest gratitude to my colleagues at the University of Texas Department of Government, who had the foresight to appoint me to the faculty before the completion of this thesis. I thank them not only for the appointment, but also for contributing to a
unique academic environment -- one that encourages a diversity of analytical approaches and strives for individual and collective excellence in both teaching and research. I particularly wish to thank Charles F. Cnudde (now Dean of Social Science at Florida State University), Thomas Ferguson, and Benjamin I. Page for their unflagging support and encouragement.
Political affairs, whether they be affairs of state or affairs of the smallest group, are conducted in conversation. The conversation relevant to a particular political decision might be the deliberations of a committee or a piece of propaganda in a "conversation" between an elite and a mass audience. Even the complete autocrat, when deliberating over a course of action, engages in a conversation of sorts, if only with himself. Language serves as the medium of such conversations, and these discourses in a language should constitute the data, *sine qua non*, of any political science. Yet, our analytical powers with respect to the data of conversations remain notoriously weak.

We may certainly engage in the hermeneutic interpretation of textual presentations of these conversations. This in fact characterizes a great deal of what political scientists do. But, due perhaps to our human frailties, we only rarely do so in any systematic fashion. Any interpreter of a text necessarily brings to a reading an entire life history of predilections and preconceptions that, while making the text comprehensible, can at the same time befoul or distort a clear understanding of the intentions embedded within the text. Perhaps if we deploy the technological resources available in the modern age in order to assist us — surely not replace us — in accomplishing interpretive tasks, we as political scientists can perform our analytic function more successfully.
This thesis is directed toward that end. It concerns the development of hermeneutically grounded techniques for the construction and analysis of computer models of political texts. It describes part of an ongoing research effort conducted in collaboration with John C. Mallery, also of the MIT Political Science Department, and it discusses some of the issues involved in grounding the system, particularly its syntactic components.

This interest in modeling the contents of political texts forces me to cross disciplinary boundaries, especially into the subject domains of linguistics and artificial intelligence. Some political scientists and linguists occasionally question the legitimacy of this. Researchers in artificial intelligence complain less often, presumably because they themselves perceive AI to be interdisciplinary.

The complainants in both linguistics and political science perceive my intellectual travels as some sort of transgression against unwritten but nonetheless somehow enforceable norms of disciplinary sovereignty. The tacit premise of the question holds that in order to cross this disciplinary boundary I require some sort of visa from the linguistic authorities (to be stamped in a passport granted to me, I suppose, by the appropriate political scientific authorities).

There are easy responses. When posed by linguists I immediately counter with my own question: by what authority does Noam Chomsky, arguably the premier linguist of the 20th century, cross this same boundary, but in the opposite direction? Linguists have no answer. Some political scientists, on the other hand, counter that Chomsky's boundary transgressions are just as illegitimate as mine, as if
politics were their special province. But this is too glib. A better counter -- one that answers critics from both disciplines -- that disciplinary boundaries are artificial, drawn for the administrative convenience of college deans. When elevated metaphorically to the status of political boundaries separating sovereign academic disciplines, these administrative conveniences become odious hindrances to scientific progress. Much can be gained from the cross-fertilization that interdisciplinary research engenders, and much can be lost from small-minded efforts to enforce rigid specialization. While to me this answer is eminently reasonable, it does not suffice. My interlocutors are owed an answer that clearly states my motives for crossing disciplinary boundaries, regardless of their imaginary status. Below, I discuss the reasons why the study of language is important for political science. Next, I turn to a discussion of artificial intelligence, stating why my collaborator Mallery, and I have chosen to implement our own methods for textual modeling rather than simply applying existing implementations. Finally, I provide an overview of the chapters that follow.

1.1 Language and Politics

What do language and politics have to do with one another? Quite simply, they have everything to do with one another. Dallmayr [1984] notes that, in the tradition of Western thought from Aristotle to the present, two factors distinguish *homo sapiens* from other species. These are language and politics. We are at once both *homo loquens* and *homo politicus*, and these two interdependent, essential features of our being stand at the center of our conception of human rationality.
Gadamer expresses the Aristotelian conception of the language-politics nexus in a way that highlights the dependence of *homo politicus* on his expressive capacity:

The distinguishing feature of man . . . is his superiority over what is actually present, his sense of the future. . . . Aristotle adds that with this the sense for right and wrong is given — and all because man, as an individual has the *logos*. He can think and he can speak. He can make what is not present manifest through his speaking, so that another person sees it before him. He can communicate everything that he means. Indeed, even more than this, it is by virtue of the fact he can communicate in this way that there exists in man alone common meaning, that is, common concepts, especially those through which the common life of men is possible without murder and manslaughter — in the form of social life, a political constitution, an organized division of labor. All this is involved in the simple assertion that man is a being who possesses language [Gadamer, 1966: 59-60].

Only through discourse can we provide peaceable gestures that can disclose to others our interpretations of our individual and collective histories, our present needs and desires, and our visions of possible futures. The alternative gestures — alternatives to which we too often turn — are gestures of coercive force.

Dallmayr believes the language-politics nexus to have profound meaning for normative political theory. But, as his own exegeses illustrate, more general epistemological issues mediate the nexus. Insofar as a lifetime of verbal practice molds the concepts and categories that emerge in the adult mind, language plays an essential role in the social arrangements we consider just. But the general nature of mind mediates that role. Any implications of language for normative political theory, then, are conditional upon the essential nature of the knower of the language.
The case for the interpenetration of language and politics can perhaps better be fashioned through examination of the role of language in the conduct of practical politics. Political decisions everywhere flow either from the outcome of political debate and discourse or, when incommensurabilities appear insurmountable, from its breakdown. While Dallmayr certainly recognizes this, he misses the fact that the centrality of language for political practice makes political utterances -- the political debate, the editorial, the manifesto, the political plan and counterplan, etc. -- the consummate data of political science. The language-politics nexus is as relevant to the development of positive political theory as it is to normative political theory.

1.2 Language, Politics, and Method

After having decided to cross the disciplinary boundary between political science and artificial intelligence (perhaps we crossed more than one; the two fields hardly seem contiguous), Mallery and I ventured naively to the MIT AI Laboratory seeking access to the appropriate "modules", which we would then integrate in order to produce the ultimate text analysis system for social scientific applications. The first person we stumbled across was Lisp guru Richard Greenblatt, who, upon learning of our plans, promptly handed us copies of the Lisp Machine Manual and advised us to start programming.

We soon discovered that Greenblatt's advice was appropriate. Existing AI parsing and knowledge representation systems were designed to demonstrate hypothetical cognitive processing principles. These principles may be demonstrated in exceeding small representations, or
"micro-worlds". As a result, attention is rarely paid to the
generalizability of representational schemes and computational
processes. That is, they do not "scale up" to a level that would meet
our analytic requirements. Political phenomena are generally quite
complex. Any AI system designed for the representation and analysis of
political phenomena from textual data must therefore be capable of
representing large-scale texts, not just one- or two-paragraph texts.

The inability of available AI systems to support the analytical
tasks we had in mind forced us to follow Greenblatt's advice
scrupulously. We began to implement our own system, which we named
Relatus (Latin for "relation"). The implementation strictly adheres to
the principle of hermeneutic interpretation, discussed in Chapter 2, as
well as to certain other linguistic and interpretive principles
discussed in Appendix I. Computationally, we have endeavored to
support the representation and analysis of large-scale texts.

Our approach, as discussed in Appendix I, is therefore broad and
shallow. This means that we endeavor first to model a broad range of
texts with shallow interpretations. Many existing AI systems reflect
the alternative, narrow but deep strategy. In Dyer's [1983] BORIS
system, for instance, a great deal of attention is paid to the semantic
interpretation of tokens that appear in his demonstration text. The
information necessary for this interpretation is represented as sets of
hand-coded demons on the lexicon entries for the approximately 300
words in the two short texts he uses to demonstrate the system. These
encodings allow BORIS to draw all sorts of inferences about the texts,
providing a deep interpretation. But the system is narrow in scope.
It works only for the demonstration texts. In order to interpret other
texts, demons must be defined on the lexicon entries of its terms. In
order to interpret any text regardless of its content, then, demons
must be hand-coded to handle the usage contingencies of every sense of
every term in the language. Therefore, this "narrow but deep" approach
will not generate tools for the analysis of lengthy, politically
relevant texts.

Of course, this was not Dyer's purpose in implementing BORIS. It
was designed to illustrate certain cognitive hypotheses and to generate
others [Dyer, 1983: 354-356]. Our purposes, on the other hand, being
somewhat different, require a much different implementational strategy
— one that is broad but shallow. The approach is broad in that a wide
range of sentences may be successfully processed. It is shallow in
that the interpretation of those sentences, initially at least, is not
deep. We represent textual contents at an eidetic\textsuperscript{1} or phenomenal
level, gradually deepening the analysis over the course of the ongoing
implementational process. The plan for effecting this gradual
deepening is outlined in Appendix I. For a fuller discussion, one must
await Mallery's thesis.

We create an eidetic representation of a sentence by performing a
syntactic analysis and then mapping the lexical contents of that
analysis into a knowledge representation system or semantic network.
The knowledge representation system we use is Mallery's [1987]
Gnoscore, which relates tokens to each other in a way that maximizes

\textsuperscript{1}According to Webster's, "eidetic" means "marked by or involving
extraordinarily accurate and vivid recall, especially of visual images.
the efficiency of procedures that reference graph configurations of relationships. For example, consider sentence (1).

(1) The administration sold arms to Iran and gave the Contras the profits.

The eidetic represented for this sentence is displayed schematically in Figure 1-1

![Diagram](image)

**Figure 1-1:** Eidetic Representation of an Example Sentence

Certain information is deleted from this schematic presentation of the eidetic representation, such as links between the tokens and nodes representing the universal concepts of which they are particulars, tense and aspect information, and certain other bookkeeping information. Notice that the subject of both sides of the conjunction is administration-l. This means that they are one and
the same administration. The relations to-1 and to-2, however, are different instances of the "to" concept, so they have different numbers to indicate that fact.

Notice that the graph does not represent the fact that the profits were derived from the arms sale. The program does not make this inference. While we plan to add facilities to automate this sort of inference, we chose first to concentrate on representing the literal and explicit contents of texts in an efficient and consistent manner. Later procedures might infer from the meanings of "profits" and "sell" that these profits were derived from that sale. Until then, a Relatus can simply state after sentence (1), sentence (2).

(2) The profits were derived from the arms sale.

The needed link between profit-1 and sell-1 would thereby be made.²

The Gnoscore class of knowledge representation system extends earlier frame systems, like FRL (Frame Representation Language) [Roberts and Goldstein, 1977], which implements ideas suggested by Minsky [1975] for the representation of knowledge. The extensions include the instantiation of each frame as a message-passing agent, bi-directional pointers, and a system of token typing [Mallery, 1987]. These innovations all support the referential determinism of Relatus. That is, the process that merges the lexical contents of sentences into a knowledge representation does so in a way that maximizes referential

²The lexicon knows the verb forms of nouns and the noun forms of verbs, so the reference mechanism can identify the expression "arms sale" with the earlier expression "the administration sold arms".
efficiency by eliminating unnecessary backtracking in the knowledge representation.\(^3\)

As in Winston's [1982] natural language application of FRL, Gnoscere creates taxonomic hierarchies that are integrated with the token-typing.\(^4\) While not shown in Figure 1-1, each token is related to its universal token type. For example, "administration-l" is related to "administration" through a universal-particular link. (3) All administrations are organizations. If one referenced sentence (3), a universal-particular relation would be created to link "administration" to "organization".

(4) All organizations are groups. Likewise, (4) would create such a link between "organization" and "group". Thus, from the output of the Relatus parser, Gnoscere integrates an eidetic representation of sentential contents with a taxonomic hierarchy.

This sort of knowledge representation scheme is not limited to the eidetic representation of sentences. Mallery [1987] has used Gnoscere to represent relationships in high-level vision. Also, extending FRL to incorporate some of the ideas of Gnoscere, Connell and Brady [1985] have developed a system that learns two-dimensional shape descriptions

\(^3\) For a discussion of determinism in syntactic parsing, see [Marcus, 1980]. Duffy and Mallery [1984] argue that referential determinism constrains the range of plausible parsing models, possibly excluding those that do not create syntactic deep structures. Marcus could not apprehend this, because his parser was not integrated with a knowledge representation system.

\(^4\) These issues are discussed more fully in [Mallery, 1987].
from examples. FRL and Gnosere class knowledge representation systems, then, seem quite useful for representing diverse sorts of knowledge and can be made to support inductive knowledge acquisition tasks.

The Relatus Parser, described in Chapter 3, was designed to ease the mapping of sentence contents into such a representation scheme. Mallery's [1987] constraint and reference mechanisms encode the syntactic relations detected by the parser in Gnosere. At present, the system requires users to state literally and explicitly in English the relationships they want to construct in Gnosere. Subsequent semantic interpretations of sentential contents will then involve successive annotations of the eidetic representation, as discussed in Appendix I.

We have chosen this research strategy precisely because we are serious about the endeavor. We certainly could have produced a program that constructs a deep analysis of a political text, but since such a program would work on that text alone, it would not constitute a general environment for the analysis of textual contents. This implementational decision does not imply that politically relevant textual analysis will be not emerge from the project for a long time. Quite the contrary, certain sorts of analysis can emerge earlier than others, as discussed in Section I.5.2 of Appendix I. In fact, analyses in Relatus of intentionally sensitive protocols of sequential prisoners' dilemma games have already been conducted [Hurwitz, et al., 1987].
The foregoing, then, in no way constitutes an apology for the absence of a political application of the research reported here. Any such application would be premature at best and fraudulent at worst. We believe we are engaged in basic research in political science. We would have been happy simply to apply AI research to politically relevant problems, but as such research had not emerged from the laboratory, we instead entered the laboratory, rolled up our sleeves, and proceeded to do the work. This thesis presents a portion of that ongoing research.

**Thesis Overview**

Readers who are unfamiliar with Relatus would most profitably begin by reading Appendix I, written by Mallory and myself, which describes Relatus. We prepared that essay for presentation to the International Studies Association in March of 1986, so it is quite accessible to social scientists. A more technical version is under preparation and will appear elsewhere.

Chapter 2 recounts the historical development of content analysis, delineating the difficulties that have emerged since the early studies conducted during World War II. In a sense, the chapter represents my personal motivations for embarking upon this research program. As an undergraduate, I was awarded a grant to perform a content analysis of the major network evening newscasts, several daily newspapers, and National Public Radio's *All Things Considered* for a week during 1978. As a result of this effort, the interpretive limitations of content analytic techniques were made quite plain to me. The chapter concludes with a plea for a "computational hermeneutic" approach to content analysis — an approach advanced over the past two decades by Alker and
his associates [Alker, 1971; Alker and Christensen, 1972; Alker, Lehnert, and Schneider, 1985] and one that Mallery and I are pursuing with Relatus. The analysis of the content of unrestricted texts, it should be noted, stands as the ultimate goal of our research. As such, it remains quite far off. However, in the context of achieving that goal, several other sorts of analysis will become feasible. These are discussed in Chapter 2.

Chapter 3 presents the implemented design of the Relatus Parser. The parser converts sentence strings into canonical, deep-structure representations of sentential contents. These contents are thereby prepared for further analyses and mapping into a semantic network representation of textual data. The parser also has the ability to map its structures into a semantic network representation at various critical transition points in a parse. Once the learning machinery of the network representation is fully installed, we hope to simulate the induction of syntactic categories and procedures. This chapter also presents the details of the categorial disambiguator -- a component of the parser that improves its performance considerably. A section of the chapter describes several of the graph structures the parser produces.

Chapter 4 presents the theoretical motivation behind the parser's implementational strategy. Although the parser is generally transformation, it bucks many recent trends in the transformational theory of syntax. Transformation theory is reviewed in Chapter 4, and its weaknesses as an ostensibly cognitive theory are brought to the foreground. In particular, the chapter discusses the simplifying
idealizations of the language acquisition situation presumed in this theory. These idealizations so fundamentally distort the facts of language development that the generative or transformational theory cannot be taken seriously as a theory of language learning. Since its purported explanations of language acquisition ground the claims made for it as a cognitive theory, it cannot be taken seriously as a theory about how people actually do learn and process language.

Chapters 5 and 6 proceed from this point to consider the appropriateness of frame-based AI knowledge representation techniques not only for modeling textual data but also for learn how to model textual data. Chapter 5 assesses three approaches to concepts and categories, defending a contemporary version of the classical view of categories against an important modern objection. Chapter 6 shows how the classical view supports the formation of novel hypotheses. The capacity to form novel hypotheses would be helpful in any social scientific application of AI technology, but it would be crucial to any simulation of language learning. That the classical view supports hypothesis formation, novel and otherwise, constitutes strong evidence for the approach.

Chapters 7 and 8 describe two experimental components of Relatus — a General, Extensible, Lexicon Shell and a temporal indexation mechanism. Both components are necessary to support the representation and analysis of large textual corpora.

Chapter 7 describes the prototype GEL Shell — A Generic, Extensible, Lexicon Shell. This system, which exists but has not yet been integrated into Relatus, will prove central for performing the
semantic interpretations as described in Appendix I. It will encache
the constraints needed to interpret the sense of particular lexical
items. The GEL Shell is designed to operate independent of Relatus,
and thus may be used for other purposes). It does not presume that it
will be used for any particular language and is sufficiently general to
find application in natural language systems no matter what theoretical
commitments an implementor may have previously made.

Chapter 8 describes a prototype temporal indexation mechanism. It
is designed to extract temporal portions of knowledge representations
or to apply an operation to all events represented that fall within a
specified temporal interval. It finds temporal intervals within a time
range in (computational) time proportional to the logarithm of the
number of temporal intervals represented plus the number of nodes
extracted. It is implemented as a special class of binary tree, the 2D
tree. The logarithmic time complexity for extraction of nodes is
guaranteed by an algorithm for balancing the tree in logarithmic time.
The balance algorithm is described in detail, and the code for the
prototype is presented in Appendix II.

Finally, Chapter 9 concludes by assessing what has been learned in
this exercise.
Chapter 2

The Validity Problem in Content Analysis

2.1 Overview

In Harold Lasswell's famous formulation, political analysts monitor the ebb and flow of social values, such as safety, income, and deference. By definition, those who enjoy larger shares of these values are termed the political elite. Elites, according to Lasswell, preserve their ascendancy by manipulating symbols, controlling supplies, and applying violence [Lasswell, 1935: 3].

Elites and elite techniques for exercising and maintaining dominance hardly exhaust the scope of political analysis. They do, however, constitute central phenomena that political theory must explain. Quantitative methods serve well for modeling and understanding the control of supplies and applications of violence. But symbol manipulation has by and large escaped our methodological grasp.

Because we lack tools adequate for the rigorous examination of the contents of political communication, our most sophisticated models cannot or can only inadequately incorporate the informational aspects of political events and processes. From the mass perspective, these are effects due to variations in the contents of information available to alternative publics. From the elite perspective, this deficiency prevents us from studying systematically the intentions of elite

---

5Lasswell himself was by no means as hard-nosed as this formulation might imply. In less rigorous moments, he advocated content analysis as a means for making people "fraternally intelligible to one another" and providing an opportunity for "taking the other fellow into account" [Lasswell, 1942: 27].
actors. From any analytic vantage point, though, the deficiency blurs and perhaps distorts our vision. It delimits critical access to a central aspect of what constitutes the political. \textsuperscript{6} Groups are unequally able to bear the costs of acquiring of political information [Downs, 1957]. People cannot possibly process all the information at their disposal [Simon, 1979]. Social inequities and differential patterns of development can themselves systematically distort the interpretation of political information [Habermas, 1976]. Because patterns of development to some degree reflect the influence of political events and processes, any effects amplify. Yet, because we lack indicators of the contents of communications, we cannot incorporate these factors into our models. Instead, we tend to construct models that assume that political actors possess full information concerning the causes and consequences of political action.

Content analysis is occasionally applied as an instrument for measuring the contents of political communications. In general, content analysts examine the contents of political documents, categorizing the documents or portions of the documents into predefined content categories, usually performing standard statistical analyses on the resulting content classifications. While there have been some notable successes, some of which are recounted below, these procedures leave much to be desired. Anticipating issues to be discussed later, two deficiencies of content analysis stand out in particular.

\textsuperscript{6} I apply broad scope to the political here. It ranges not only over phenomena pertaining to the administration of states and the competition for state power, but also pertaining to the administration of and competition for power in any and all human organizations.
1. For results to be meaningful, content categories must be valid; the classification procedures must somehow organize textual data so that results faithfully reflect the meaning of the text under analysis. Thus, content categorization necessarily involves reasoning about the meanings of texts. A valid content categorization, on this view, would reflect a valid interpretation of a text. However, interpretations of texts may vary across individuals, depending upon the historical embeddedness or life-world of the text-interpreter. In principle, no objective decision criterion exists with which we can demarcate valid from invalid interpretations.

2. Human-coded content analyses are labor-intensive. As the size of textual corpora to be analyzed increases, labor costs quickly become prohibitive. Efforts to reduce these costs by applying computer technology to content-analytic problems have been somewhat successful, but in the process much analytic power has been lost. Because existing computational content-analytic methods interpret texts without adequately, taking into account the effects of the linguistic structures within which textual contents are embedded, the validity of such analyses is threatened.

In this Chapter, I recount how validity problems have surfaced in studies of the contents of political propaganda. I advocate "computational hermeneutics" as an alternative strategy. I distinguish my usage of this coinage from two recent usages that each, for different reasons, express the notion of computational hermeneutics inadequately. I distinguish it from a third usage which only partly expresses the implications of a marriage between computer technology and hermeneutic interpretation. I distinguish it further from the expressly non-hermeneutical approaches adopted in much of the available AI research in computational text analysis.

Computational hermeneutics, as conceived here, accepts as valid only those procedures for computerized content analysis that do not force textual data artificially into any single interpretive mold.
Instead, valid procedures are those that interpret texts according to variable parameters constituting the interpretive presuppositions of computer models of text interpreters. As parameters, these interpretive presuppositions may be manipulated systematically, thereby providing an opportunity for tests of hypotheses concerning the interpretation of textual contents under differing interpretive regimes. To see why content analysis requires this interpretive turn, we must digress to a discussion of the threats to validity it has faced in its historical development.

2.2 Precomputational Content Analysis

Harold Lasswell inaugurated modern political content analysis in his doctoral study of the content and procedures of political propaganda during the First World War [Lasswell, 1927]. His analyses, illuminating though they were, lacked methodological rigor. They were simply Lasswell's common-sense reflections on propagandistic texts, their producers, and their intended audiences.

In the introduction to the 1971 republication of his dissertation, Lasswell [1927: ix-xvii] characterized the study as an early, modest contribution to a research program that would guide his entire career. He sought to articulate a theoretical scheme for the study of "international attitudes". Lasswell believed the symbolic aspects of international conflict to be systematically understudied, and, consequently, systematically undervalued in efforts to resolve international insecurity dilemmas.

During the Second World War, Lasswell collaborated with Nathan
Leites and several junior scholars\textsuperscript{7} in an effort sponsored by the Library of Congress to study more systematically and rigorously the contents of Nazi, Soviet, and American propaganda.\textsuperscript{8} Continuing the research program begun with his dissertation, Lasswell bemoaned what he perceived to be the prevalent tendency to rely on material factors in explanations of political phenomena. He cautioned that cultural explanations seem less plausible only because we lack a methodological handle with which to grasp them.

So far as the material dimensions are concerned, operational methods have been worked out to describe them; not so with the ideological. We are amply equipped to describe such "material" changes as fluctuations in output or amount of machinery employed in production; but we can not match this part of the description with equally precise ways of describing the ideological. The result is that the historical and social sciences have been making comparisons between patterns, only a few features of which are handled with precision. The other dimensions remain wholly qualitative, impressionistic, and conjectural [Lasswell, 1949: 47].

Apart from their very real, pragmatic concerns, Lasswell, Leites, and associates sought to rehabilitate ideological or cultural explanation by liberating it from its methodological gulag.

Their rehabilitation strategy -- quantitative content analysis -- was quite straightforward. They devised content-analytic categories

\textsuperscript{7}They were Raymond Fadner, Joseph M. Goldsen, Alan Grey, Irving L. Janis, Abraham Kaplan, David Kaplan, Alexander Mintz, Ithiel de Sola Pool, and Sergius Yakobson. Many later became leading social scientists.

\textsuperscript{8}Nazi and Soviet propaganda were analyzed as part of surveillance efforts directed at an enemy and a potential enemy. American propaganda was analyzed in order to detect home-front fifth columnists.
into which coders classified specific significations\(^9\) appearing in political texts. They then related the frequencies in these categories to the pragmatic or political action consequences resulting from exposure to these significations. The resulting story would hopefully proffer a convincing account of propagandistic, symbol-manipulating efforts.

By no means did Lasswell et al. naively assume the validity of this approach. They instead pushed the problem of validating the content-analytic categories to the foreground. Janis [1949] discussed the critical threat to the validity of the approach. The operation of classifying textual fragments into content-analytic categories requires some assessment of the meanings of those fragments. Not only might the investigator's assessments of the meaning of any text diverge from those of its producer and audience, but the meanings assigned by the producer and audience may even diverge from one another. Moreover, terms may signify differently for different members of an audience or by different subgroups within an audience.

Content analysis lacked a secure Archimedean point. The belligerent political context of the research effort limited the extent to which Lasswell and his associates could engage in validation studies. The producers and intended audience of the texts under analysis were, after all, citizens of nations at war (or close to war) with the nation in which and for which the researchers worked. The collection of criterion measures against which to assess the validity

\(^9\)Signification refers, simply enough, to the things to which a sign or symbol refers [Morris, 1964: 3].
of content categories could not have been more unfeasible. So Janis
innovated, proposing an "indirect" method of validation.

In general, the more closely the relationship [between a
content variable and a variable representing subsequent
political action] approximates that of quantitative
covariation between the content characteristic and some
pragmatical characteristic which is most directly tied up
with semantical signification responses, the higher the
probability that the content characteristics which enter into
that relationship are validly measured. This conclusion is
based on the assumption that validity implies a one-to-one
correspondence between individual signs and the signification
responses of sign-interpreters. The relationship which most
closely approximates testing the one-to-one correspondence.
.. has the greatest weight as evidence of validity [Janis,
1949: 79-80].

Valid procedures would be those that produced content characteristics
that correlated with subsequent political actions by the people
(sign-interpreters) who had been exposed to those contents.

Janis recognized that his indirect technique could not
sufficiently validate any particular content categorization scheme. The
technique fails to assure us that negative content-analytic results do
not increase the likelihood that a particular hypothesis is false. As
Berelson [1952: 171] explained much more clearly, low correlations
might signify low relationships but high validity of the content
categorization scheme. Alternatively, they might signify high
relationships with low validity. For this reason, Janis cautioned
against the use of content analytic procedures to disconfirm
hypotheses. While this imposed stringent limits on the usefulness of
quantitative content analysis as a tool for theory development, it
posed no special limitation for serving the practical end of the
analysis, service to the military command.
As it turned out, the Library of Congress study did not succeed in serving its war-related ends. As participant Ithiel Pool recounted some years later,

As an intelligence operation it [the Library of Congress study] was not particularly successful. It was operating beyond the state of the art, and as a result contributed far more to content analysis methodology that might at some future time contribute to intelligence, than it contributed to substantive understanding of the enemy at the time [Pool, 1967: 15].

Whatever its limitations as a tool for hypothesis testing, quantitative content analysis might be useful as a descriptive tool for the study of political propaganda. Given their practical aims, and despite the limited validity of their procedures, the content-analytic studies of Lasswell, Leites, and their associates succeeded to the extent that they were indeed able to generate such descriptions, even if they arrived too late to aid the war effort.

We may attribute their success to the fact that human coders interpreted the texts holistically. The applications presented in *Language of Politics* each focus upon exceedingly high linguistic levels. Units of analysis are the text as a whole, the headline, and the slogan. Inter-coder reliabilities for categorizations of the contents of headlines exceeded 0.9 when the categories were pre-defined by the investigators [Kaplan and Goldsen, 1949]. Human coders, then, were apparently able to place texts into thematic categories relatively reliably when texts were considered as wholes.

While Lasswell, Leites, and associates may well have been willing to accept the limitations on the validity of their approach, foregoing
theory-building in favor of more practical pursuits, maintenance of a high degree of reliability was crucial. Unreliable results would subvert even their practical aims. To maintain accurate results, human coders required extensive training, and multiple coders were needed in order to detect deviant and possibly mistaken codings.

Coverage, too, was important. Random samples of propaganda could certainly be collected and analyzed, but random selections of propaganda artifacts would satisfy only the theory-formation purposes that had already been foregone. The practical aims of the research demanded more complete coverage of the data. Yet there was a wealth of data -- too much data, in fact. If the data coverage was to be adequate for the practical aims of propaganda analysis, the labor costs for the needed multitude of human coders would prove prohibitive.

This must have presented something of a dilemma, and we can at least surmise that the labor-intensiveness of the endeavor contributed to the failure of its practical aims. The use of multiple, trained human coders limits the effective scope of content analysis. Large corpora cannot feasibly be analyzed for content, due to the high labor costs involved. Likewise, investigators seeking to analyze texts more thoroughly, delving beyond headlines and into the text must also be willing to pay these costs. If the new computing machines could be used to process and analyze textual data, it was thought, the effective scope of content analysis might be broadened.

2.3 Early Computational Content Analysis

With his General Inquirer, Philip Stone offered hope that this dilemma might be resolved. The General Inquirer scans textual input
and automatically classes words into content-analytic categories or "tags" defined by a lexical database or dictionary [Stone, et al., 1966]. Lasswell, already a booster of social scientific applications of the new computer technologies [Lasswell, 1951: 141], saw the General Inquirer as the potential salvation for content analysis.

The emergence of the computer has made it feasible for social and behavioral scientists to make a fresh start on content analysis. The vast potentialities of content analysis, though foreseen for some years, have been poorly realized, owing chiefly to the onerous task of scanning texts and processing data. . . .

The essential problem of content analysis is to provide a procedure that preserves and codes as economically as possible the meanings relevant to social theories. The scientist who works with samples of data with which he is concerned can now hope eventually to identify the universe of referents pertinent to his theory. His instruments will then faithfully present the distribution of relevant meanings intended or perceived by the participants in the social process. In the universalizing civilization of tomorrow, taking samples of communication at strategic spots will, when properly harnessed in an inclusive man-machine network, keep signs and meanings in very close harmony with one another. Or if not, the discrepancies will reflect the vagaries of policy rather than the dimness of incapacity [Lasswell, 1966: viii-ix].

Applications of the General Inquirer have produced some interesting results. Apart from the early results reported in the volume by Stone et al. [1966], Namenwirth's [1973] study of Democratic and Republican party platforms and Weber's [1978] study of British Speeches from the Throne stand out as two exemplary quantitative content-analytic applications with political relevance. Using the content categories of Lasswell's Value Dictionary [Lasswell and Namenwirth, 1968], these studies detected cycles in the value concerns of the elites who produced the analyzed documents. The Namenwirth and
become extinct. If any feature of human languages may be said to be
universal, surely its openness to human creativity must be one of them.
The meanings of the word-tokens classified into the rigid category
scheme prescribed in the Lasswell Value Dictionary may certainly have
undergone subtle changes across the 120-year scope of Namenwirth's
study and the 283-year scope of Weber's study. 10

Pursuing Namenwirth's suggestion, Weber [1987] reports results of
comparisons of the value cycles detected in his study 11 of British
Speeches from the Throne to economic and demographic time-series and to
predictions generated from a widely-believed sociological theory.
Happily, Weber presents excerpts from the actual texts that indicate
that diachronic changes in meaning do not substantially affect the
results. He conducts statistical tests of the following propositions:

1. Cycles in expressed values reflect cycles in macroeconomic
   performance.

2. Cycles in expressed values reflect the generational
   population replacement or cohort succession.

10 Of course, word senses may even vary internally. A particular
word, for instance "state", may have many different usages. Kelly and
Stone [1975] have developed procedures, incorporated into the General
Inquirer, which disambiguate word senses on the basis of neighboring
lexical items in the text. These procedures examine the stream of
words, however, and not the structural relationships between words and
sentences. Thus, they cannot perform as well as human word-sense
disambigulators, who have access to the syntactic, semantic, and
pragmatic structures underlying the surface text. While these
procedures reduce word-sense ambiguity as a threat to validity, they do
not eliminate it. More cognitively defensible approaches, based upon
the positions of ambiguous terms in syntactic and semantic structures,
have been presented by Small [1980] and Waltz and Pollack [1984].

11 Weber uses the Lasswell Value Dictionary, further classified
into broad thematic categories on the basis of prior factor analyses of
the value tag assignments in the texts.
3. Cycles in expressed values are predicted by the Bales-Parsons Theory of Social Action.

Weber is unable to demonstrate or reject any of these propositions conclusively. However, his results are suggestive, warranting further study. At the very least, this research suggests that computerized content analysis may indeed play a role in the construction and evaluation of macrosociological theory.

2.4 The Limits of Word-Frequency Content Analysis

Namewirth and Weber are able to generate suggestive results because their research examines texts at a macrosociological level of analysis, and because they limited their research interest to the broad, thematic content of their textual data. But this does not begin to exhaust the scope to which content analysis might be applied. Many scientific applications of computational content analysis can be conceived at lower levels of analysis. These applications might require inferences not only about the broad, thematic contents of texts, but the directionality of opinion and the implications of propositions as well. For example, Bales has for years wanted to apply such techniques to the verbal protocols extracted from his SYMLOG group psychotherapy sessions [Bales and Cohen, 1979: xiv].

---

As an initial application of the General Inquirer, Bales wanted to use it to provide real-time feedback to participants engaged in group psychotherapeutic interaction. This goal, write Bales and Cohen, has never been achieved. The obstacle, according to Stone [personal communication] has been at the data entry stage. Group interaction occurs too fast to allow real-time entry of the contents of verbal exchanges. For the SYMLOG sessions, Bales is interested in scoring the level of imagery e.g., personal, social, fantasy) and the affective direction (dominant-submissive, friendly-unfriendly, instrumental-emotional) of utterances in the session. The General Inquirer does
analytic techniques have even been applied provisionally at the individual level of analysis to verbal data extracted from applications of projective psychological instruments, including the Rorschach Test and the Thematic Apperception Test [Viney, 1983].

Let us consider a hypothetical example to show that while the General Inquirer can generate interesting, and possibly valid, conclusions at the macrosociological level, it fails at lower levels of analysis. Using the Lasswell Value Dictionary, the General Inquirer would categorize the following sentences equivalently:

1. We should raise taxes.
2. We should lower taxes.

Both sentences would accurately signal an instance of the content category "concern for wealth". A political campaign in which two candidates adopted these opposing positions and campaigned on them strenuously would indeed be one characterized by concern for issues of wealth. The General Inquirer can detect this. However, when the research interest moves from this macro level to more micro levels of analysis, a central deficiency becomes painfully obvious.

Suppose, for example, that the research goal is to characterize the nature of the political opposition in such campaigns. The General Inquirer can tell us nothing of the directionality of the value concerns. As a result, it can detect no difference between the issue seem capable of reliably and validly assessing the level of imagery, but, since they involve complex interpretive inferences, it is doubtful that the General Inquirer could produce reliable and valid scores for the affective direction of utterances.
positions adopted by each candidate. Legitimate applications of the technique are limited, therefore, to studies, such as those of Namewirth and Weber, that seek to characterize broad value concerns over large textual corpora and that refrain from offering conclusions regarding the directionality of those concerns.

Even at the macrosociological level, some degree of skepticism remains warranted regarding the presumption that raw word counts can accurately represent the value concerns thematized in text. In very large textual corpora deviations between textual meanings and General Inquirer category assignments might wash out, but this is not necessarily the case. The variations thus produced might just as easily be systematic. More importantly, the invocation of one theme, strategically placed in only one sentence of a text, might tie together whole sections of the text, providing it with coherence. Because it ignores the structure of the text, computer-based, quantitative content analysis would weight that single invocation equally with all other value invocations, even though without that invocation the text as a whole might lose its clarity, its comprehensibility, and its capacity to mobilize or demobilize political action.

The deficiency of the word as unit of analysis emerged as a central theme in the papers presented at the 1955 Allerton House conference on content analysis [Pool, 1959]. Ithiel Pool discussed the significance of this theme retrospectively. He wrote that the appropriate unit depends upon the purpose of the research. A study of pausal phenomena in speech, for example, would take the phoneme as the appropriate unit of analysis. For content analysis, however, we
require identification of a basic unit of meaning. For propaganda analysis, that unit might be the slogan. For analyses of plot, the incident of action might be most appropriate.

We are here implying an important point made by a number of the authors of this volume: categories of analysis should be related to the structure of the material under discussion. [Robert Plant] Armstrong makes this point about literary materials, objecting to some earlier content analyses of literary materials in which the items counted had little to do with literature as such. [Alexander] George makes this point about propaganda analysis; he finds that symbol counts which took no account of the structure of strategic planning in the propaganda yielded few inferences of use. . . . [Sol] Saporta and [Thomas] Sebeok make the same point with regard to such material as charms, riddles, portents, etc., which conform to strict formal rules; the appropriate unit to use in such cases is a unit of the structure.

Since most inferences are based on which of alternative symbolic forms actually appear, and since structural constraints partly determine the range of alternative possible forms, it is clear that for most purposes what one counts needs to have some relation to the structure of the piece examined. . . . A mode of analysis which does not permit one to take account of . . . structural considerations but requires the treatment of statements on all points as alike is likely to be a failure [Pool, 1959a: 204-205].

Pool noted that a more linguistic approach to content analysis was necessary if these structural considerations were to be taken into account, and he recommended Zellig Harris's research program in structural linguistics [Harris, 1951] as a useful starting point for content analysts who wished to become more linguistically sophisticated in order to avoid "the dangers of mechanical absurdity" in content analyses based solely on word frequencies [Pool, 1959: 227].

One of the co-authors of The General Inquirer encountered this same problem, commenting that
When we rely on the tag tally [frequencies] as our sole source of information, we risk the possibility of entirely missing the meaning of the tag distributions. Just because the count is high on a particular tag does not mean that we understand what it means. Perhaps the only solution to this problem is to... make retrievals on the basis of that tag word and study the sentences extracted from the documents [Ogilvie, 1966: 463].

While perfectly reasonable, this solution would minimize the labor-saving impact Lasswell had hoped the General Inquirer would provide.

The program's blind pigeon-holing of word-tokens into "tags" or content categories itself represents a threat to validity. Only when a human interpreter actually reviews the documents, as did the coders in the Lasswell and Leites studies, are valid content categorizations likely to emerge. The program's blindness to the meaning of the text follows directly from its inability to model the syntactic, semantic, and pragmatic structures essential to all linguistic communication. Word-frequency categorization techniques simply do not conform to these widely recognized constructs of linguistic theory. When these structures are omitted from the accounting, content categorizations of word-tokens interpret linguistic data in ways which violate the nature of language itself. The underlying, unspoken, and patently absurd assumption of computerized, quantitative content analyses is that the whole (the text) is equal to the sum of its parts (the frequency distribution of words).

13Linguists disagree over the appropriate form of these constructs, but there is certainly a general consensus that natural language cannot be understood when surface streams of word-tokens are considered in isolation from their syntactic, semantic, and pragmatic relationships.
Summarizing in programmatic outline the best speculation of linguists and philosophers of language, David Hays told a 1967 gathering of content analysts that:

Every person has a store of knowledge, not in the form of sentences, but in the form of a network of things, actions, and properties. Understanding a message consists in transducing it into a form comparable to that of the stored network, and making tests against what is known. In some cases, the new statements thus obtained are added to the network of permanent knowledge [Hays, 1969: 65].

Existing techniques for computerized quantitative content analysis do not simulate the process of human text understanding. The techniques transform the data into unnatural categories, and this constitutes a serious threat to the validity of content analyses of this type. The procedures themselves pollute the construct validity of any substantive results.14

Weber (1985: 37-38) recommends the determination of category schemes by analysis of the covariation of tokens in the textual corpora to be analyzed. While this procedure grants to the content analyst some assurance that the analytic categories are not entirely ad hoc, it does not provide construct validity. The linguistic relationships that differentiate texts from mere lists of words constitute essential elements that cannot be eliminated from any analysis without posing a threat to validity. Likewise, contextual features, including not only the historical context in which the text was produced, but also the communicative context between the text producer and text recipient, simply cannot be discarded as irrelevant to the analysis.

14 For a discussion of construct validity in the context of content analysis, see [Krippendorf, 1980: 167-168].
Concurrent with the Library of Congress study of Laswell, Leites, et al., Hans Speier [1941] directed a group for the Foreign Broadcast Intelligence Service of the Federal Communication Commission. The FCC group analyzed the contents of German radio propaganda. While in the Library of Congress study inferences concerning Nazi strategy intentions were based upon statistical tests of the quantitative content categorizations of human coders, the FCC study was far more qualitative and impressionistic. Years after the war, Alexander George, a junior participant in the FCC study, evaluated the results of the research, comparing the inferences made during wartime to the relevant historical record that had by then become available. He discovered that approximately 80% of their inferences were correct [George, 1959: 260-269]. Pool, a participant in the Library of Congress effort, agreed in retrospect with George that the FCC study was far more successful [Pool, 1967: 22].

George [1959: 96-105] attributed the comparative success of the FCC study to the greater flexibility of its qualitative approach. Pool responded that, although the quantitative techniques employed in the Library of Congress study were indeed far less successful than the qualitative techniques employed in the FCC study, the computerized variant of quantitative content analysis might still serve as a useful tool for the qualitative analyst.

Alex George was pitting a clumsy and now outmoded content analysis technique against the intuitive skill of the analyst. . . . [C]ontent analysis as George evaluated it required that the analysis system and categories had to be fixed in advance and carried on uniformly regardless of the question put by the analyst. It was this flexible adaptability to the needs of a complex inference structure
that George identified as the weakness of quantitative content analysis in World War II intelligence. A modern computerized retrieval and analysis system, however, would have all the texts in bulk memory, and would produce rapid scanning and analysis of them after the analyst had come up with a question generated by his intuitive processes. The question and the mode of analysis would be specific and pinpointed to a particular phenomenon that had attracted the analyst's attention. . . . Full time analysts are pretty good at . . . scanning [large text corpora]. Aided by a computational retrieval and analysis system they could be better [Pool, 1967: 23].

The General Inquirer would constitute the core of the computational workstation Pool proposed. As we know from his commentary on the 1955 Allerton Conference, excerpted above, Pool was quite aware of the limitations of the word-frequency basis of the General Inquirer. Nevertheless, Pool felt that the computational approach might help make more manageable the vast amount of unprocessed textual material produced each day.

In the 20 years since Pool broached this workstation concept, technical advances in computation have made possible the development of workstation environments that can provide contextually-sensitive textual models that simulate the process of human text interpretation. The development of symbol-manipulating computer systems and languages and modern advances in our ability to process and represent knowledge from textual sources provide may of the technical prerequisites.
2.5 Natural Language Understanding: AI Meets Hermeneutics

The potential utility of AI techniques for social scientific research is by now widely recognized, as a growing stream of special-purpose social scientific applications has begun to appear. Research in natural language processing (NLP), a major sub-field of AI, has progressed to the point where its application to content analysis might reasonably be considered. Certainly no one has completely reduced any natural language to an algorithm. But, since that eventuality is, at the very least, highly unlikely and arguably impossible, given the mutability of language, we need not wait for it.

Suppose we take seriously Pool's and Hays' suggestions and avoid the "mechanical absurdity" of previous computational approaches to content analysis. We might do so by applying NLP techniques to model the linguistic structures of the texts we wish to content-analyze.

Alker [1975] suggests just such a strategy as one which can produce process models that incorporate not just the causal relations between variables, but intentional ones as well. As intentions express the interests of human actors in political situations, such an approach to modeling should be seen as one that is expressly political — i.e., more than merely the (only partly satisfactory) application of natural science techniques to the human world of politics. On the crucial distinctions between the domains of the natural sciences ad the human sciences that make the development of intentional and interpretive techniques a priority, see Habermas [1971].

As a by-product of this effort, incorporation of essential linguistic structures into our analyses would improve the construct
validity of computational content analysis. AI can thus be, as Weber [1983: 128] suggests, a useful supplement for word-count approaches. However, AI-based text-interpretation will more likely supplant word-count analyses than supplement them.

Existing NLP systems can create data structures amenable for analyses of the semantic and pragmatic relationships intended by a text producer or inferred by a text recipient. Many AI researchers are actively engaged in such a research program. As a result of their efforts, we now recognize more clearly the difficulties of producing textual models which are at once both linguistically adequate and computationally tractable.

While we cannot realistically expect AI research to provide us overnight with computational tools amenable to content analytic applications, AI natural language researchers are making steady progress toward the goal of computer text understanding, simulating human performance in constructing and interpreting the syntactic, semantic, and pragmatic structures that are essential to language comprehension and production. However, some proposed AI models of human linguistic performance, because they misconstrue the meaning of interpretation, offer little or no hope of attaining that goal.

Natural language texts, unlike data matrices, can express a wide range of phenomena, encompassing propositions about causes, intentions, reasons, affects, attributions, etc., all of which have or should have import in explanations of political events and processes. A wide range of pragmatic analyses that generally fall under the rubrics of "text linguistics" and "discourse analysis" can be and should be deployed, as
Alker [1975] advises, in order to construct and critically evaluate models that can capture, reflect, or represent political phenomena in a manner consonant with our experiences as political animals. AI representational techniques can provide a useful substrate for such political analyses, but, to be useful, they must not be deployed in a way that violates our understanding of the political and our sense of political legitimacy as involving a free and open discourse among interpreters with differing interests and interpretation—a discourse that seeks to attain unconstrained consensus as the ground of truly rational action.

For the remainder of this chapter I want to articulate the general direction I believe we must should take if we hope to construct such a substrate—an environment for textual analysis that does not impose interpretations upon producers and interpreters. This would be an environment in which valid textual analyses could be conducted.

2.5.1 Hermeneutics as Validity

Let us reflect on the notion of validity. What does it mean to ascribe the predicate "valid" to a textual interpretation?

Validity assessments with regard to measurements of the meanings of texts differ radically from assessments of the validity of numerically-indexed instruments. Standardly construed, when we inquire into the validity of an instrument, we ask whether it measures whatever we intend it to measure. When we inquire into the validity of some measure of the contents of a text, however, we ask whether our measure represents the meaning which was intended by the text. In other words, we ask how the text was understood. The validity problems that
confront frequency-oriented content analyses reflect the enormous
capacity for misunderstanding embedded within those techniques.

Social scientists attuned to the fact that the living, breathing,
human objects of their sciences are themselves potential subjects of
science adopt a particular, interpretive stance toward their objects.
For the social scientist, unlike the natural scientist, meaning rarely
comes pre-packaged as a set of rigidly defined and institutionally
enforced disciplinary norms. Meaning is often subject to negotiations
between the social scientist and his or her observational object.
Understanding involves some fusion of the interpretive horizons of the
two. In other words, the social scientific observer who hopes to
comprehend his object (another subject) must make the horizons of his
life-world meet those of his object [Schutz, 1967].

Likewise, understanding a text (and thus also the validity of any
text interpretation) necessarily involves fusing the horizons of the
reader with those of the text producer, who, in producing the text,
atttempts to fuse his horizons with those of his intended audience. In
short, the validity of meaning interpretations can only be assessed
with respect to the life-experiences and historical embeddedness of all
participants in the text -- the text producer, the intended audience,
and even the (sometimes) unintended reader, the subject of science.
This is what is meant by hermeneutics.\textsuperscript{15}

\textsuperscript{15}Gadamer [1976] provides a useful introduction to hermeneutics.
For a review of competing positions within the literature on
hermeneutics and for a discussion of computational approaches to
hermeneutics, see [Mallery, Hurwitz, and Duffy, 1986, 1986a]. For
discussions of hermeneutic approaches to political science, see [Alker,
1975; Moon, 1975].
A Computational hermeneutic, then, would consist of a system for the analysis of texts that refrains from imposing interpretations upon the texts to be analyzed. Interpretations would instead by grounded in the previous interpretations of a simulated interpreter. As successive interpretations accumulated, as new textual data arrive and are incorporated into a model of the beliefs of an interpreter, the uniquely individual aspects of the interpreter's life-world begins to emerge. Other such models can be constructed for other interpreters. Differences in, for instance, the conceptual intentions across interpreters result in the accretion of partly commensurable, partly incommensurable belief sets that emerge. This produces an environment in which the symbolic aspects of conflict, for example, can be studied systematically, as Lasswell had dreamt. Interpretive presumptions can be altered and inferences compared in order to test hypotheses concerning the salience of particular presumptions for the attitudes, intentions, and what not, that emerge.

I want to advocate a computational hermeneutics that is similar in fact and exactly equivalent in spirit to the "formal hermeneutics" advocated for years by Alker and his associates [Alker, 1971, 1979; Alker and Greenberg, 1971; Alker and Christensen, 1972]. Relatus, in fact, is designed as a "framework within which the perceptions, norms, and proposals of different groups can be coded" [Alker, 1971: 35]. However, with Relatus, these perceptions, norms, proposals, etc., will not be coded by forcing them into formal structure on the basis of some prespecified set of conceptual categories. Instead, Relatus interprets the textual productions of political actors on the basis of conceptual
categories derived from the prior (and postulated) productions of the individuals and groups under analysis (see also section I.2.3 of Appendix I). Relatus, then, should be seen as more than just another AI program. It represents a hermeneutic substrate for conducting the sorts of political analysis Alker and associates have previously demonstrated to be feasible.

2.5.2 Hermeneutics versus Artificial Intelligence

For AI natural language systems to be both plausible as a models of human performance and useful for social scientific applications text interpretation cannot be based upon any one set of pre-programmed interpretations. It would instead be grounded upon knowledge already represented as a result of the system's previous interpretations. Expressed in the metaphorical language of philosophical hermeneutics, this means that the system's interpretations should depend always upon its interpretive horizons, or the contents of its beliefs about the subject matter and about the source and intended recipient of the communication, at the time of interpretation. One would be entitled to characterize such a computer model of human linguistic performance as "computational hermeneutics".

AI natural language research has, unfortunately, generally avoided the hermeneutical problematique. Most AI natural language researchers have historically preferred to reduce the lexical contents of texts to rigid, pre-defined "semantic primitives" imposed upon the system by the its implementor. Perhaps the most widely known version is Schank's and Abelson's [1977] conceptual dependency theory.
Much of value may be found in the work of Schank, Abelson, and their students, particularly at the higher and more pragmatic levels of textual analysis (e.g., analyses of narrative structures and goal structures). Insofar as their work integrates the pragmatic level of goals and plans into a computational linguistic framework, it represents an advance [Krippendorf, 1980: 128]. In this respect, aspects of their research might find fruitful application within the hermeneutic approach advocated here.\textsuperscript{16} At bottom, however, the "semantic primitives" upon which they base all interpretations are the rigid, pre-programmed interpretive assumptions of the implementors.

"Semantic primitives" are \textit{ad hoc}. No one has presented principles which might guide us in the decomposition of lexical items. Natural languages are self-referential. Any usage of a word-token may be paraphrased by (linguistically structured) sequences of other word-tokens. This being the case, we have no decision criterion by which to select the word-tokens to be understood as primitive, as opposed to the word-tokens which are to be understood as compositions of primitives [Mervis and Rosch, 1981: 104-106].\textsuperscript{17}

Actually, "semantic primitives" are not primitives at all. Schank and Abelson offer them as atomic concepts to be combined into more complex concept-molecules but they present no evidence to indicate that

\textsuperscript{16}One hopes, however, that the discourse structures (or "scripts") posited by proponents of the approach can be induced mechanically rather than pre-programmed. Kolodner [1983] appears to have made some important steps in this direction.

\textsuperscript{17}For a similar critique of a related approach advanced within generative linguistics, see [Bollinger, 1965].
concepts actually form this way. PTRANS and its friends have the status of categories, not "primitives."

Thus "semantic primitives" do not differ in status from the content categories of the Lasswell Value Dictionary. They are alleged to be exhaustive categories into which (and combinations of which) all textual contents may be pigeon-holed. The content categories implicit in this decompositionalist approach require validation, just as the categories of quantitative content analysis must be validated.

No validation studies have been forthcoming from the decompositionalists for precisely the same reason that validation eludes the quantitative content analysts. Cross-subjective validation is not possible because an interpretation depends upon the interpretive horizons of the interpreter. The interpretive presumptions of any one interpreter, be it Harold Lasswell or Roger Schank or anyone else, simply do not count as the one and only valid interpretation for all people and for all time.

Throughout the history of Western civilization, philosophers have attempted to list exhaustively all the higher-level constructs from which everything in experience can be seen as a specialization. Aristotle had ten such concepts. The Stoics had four. Locke denied the essential predication of categories, considering concepts to be conventional. Kant detected 12 pure and a priori concepts. Hegel developed his dialectical logic in an effort to show how, from Dasein (or the dialectic of Being and not-Being), a cornucopia of categories could emerge. Peirce reduced the number of categories to three — the monad, the dyad, and the polyad — each differentiated from the other
on the basis of the number of terms in a predication. Husserl, like Schank and Abelson, believed that the first task of a pure philosophical grammar is the articulation of the "primitive" forms of meaning, which can then be composed in various ways to account for the full range of expression [Thompson, 1967].

Following Frege's foundational work in set theory and Russell's development of the theory of types and its paradoxes, Wittgenstein first argued the futility of seeking to develop a universal ideal language (or a theory of categories) which can state limits of cognitive meaning.

In his early work... Wittgenstein spoke of the limits of cognitive meaning as the ineffable, as what can be shown but not said. In his later writings he repudiated the suggestion that the limits constitute an ineffable subject matter, something to be unveiled but not articulated as a theory by philosophical analysis. However, with the assumption of such subject matter philosophical clarity is to be achieved by the construction of an ideal language, a language [that] is stripped of all superfluous symbolism and is hence unable to give the illusion of transcending the ineffable limits of cognitive meaning. But if this assumption is itself an illusion, as Wittgenstein later held, if we can no more show than we can state the limits of all language, then philosophical clarity can be achieved only piecemeal, context by context; there is no short cut via an ideal language. And a fortiori there is no universal scheme of categories to be unveiled, let alone to be established by a theory [Thompson, 1967: 54].

For the later Wittgenstein, language is not static. It is a game played by everyone, the rules of which change form discourse context to discourse context. While Apel [1965] rightly observes that language has a historicity that provides it some continuity (language is a joke our ancestors play on us and we on our progeny), it certainly cannot be described in any static, putatively universal set of "primitives." language is far more fluid, and thus also far more useful, than Schank and Abelson would allow it to be.
radically simplified inferential operations in the knowledge base. If an inference engine need only deal with one or two dozen relational categories, the speed of its deliberations increases dramatically. By tossing away the subtle nuances of meaning that pertain within particular discourse contexts (language games), an unacceptable analytical risk is run. The information loss might easily result in faulty inferences.

In the lexicalist approach outlined in Chapter One and in Appendix I, we can have the best of both worlds. By relating lexical relations to other nodes that represent "conceptual primitives," the analytical advantages of the category scheme are made available while full information is retained. Moreover, because any node may be predicated any number of times, this approach does not limit the analyst to any single category scheme.

Suppose, for instance, that we represented the sentence "The Israelites fled to Sinai" and that we anted to denote the relation as an instance of PTRANS and as an instance of a category we'll call "political action." The representation appears in Figure 2-1.
The compositionality of the network structures guarantees that the generalizability criterion will be satisfied. The generalizability criterion counsels us to ensure that the network structures are flexible enough to support any sort of analytical requirement that may be imposed upon represented sentences.

In his monograph summarizing the techniques of content analysis, Weber [1985: 72-77] concluded by prescribing the adoption of a "computational hermeneutics" as a remedy for the interpretive infirmities of quantitative content analysis. But Weber's notion of computational hermeneutics is apparently as far removed from ours as conceptually possible. In fact, the model he touts as exemplary -- Dyer's [1983] BORIS system -- in fact is the latest product of the
decompositionalist school.\textsuperscript{18}

In an analysis they term "explorations in computational hermeneutics, Alker, Lehnert, and Schneider [1985] apply Lehnert's plot unit analysis technique to the events recorded in the Christian Gospel leading up to the crucifixion and resurrection. The Alker et al. work is hermeneutical in two senses. First, in seeking to extract the story's mimetic plot structure, they seek to articulate its "infectious, viral, self-replicating power," that is, its power to serve as an interpretive template for action, particularly political action, in the Christian communities subsequently infected.

The second sense in which their study was hermeneutical revolves around the process of representing the story in Lehnert's formalism, different coders constructed different representational interpretations of the actions of the story's characters. These differing interpretations reflect the differing effective histories of the coders themselves—the interpretive presuppositions that they brought to and imposed upon the text. Here is a fertile ground for research. How do different interpreters understand a text? What interpretive presuppositions account for intersubjective differences in interpretation? How do interpretive similarities and differences

\textsuperscript{18}Weber (personal communication) now agrees that the underlying assumptions of Dyer's research are fundamentally incompatible both with the notion of computational hermeneutics and with the conception of social science which he and I more or less share. At deeper (i.e., more pragmatic) levels of analysis, Dyer's work (and indeed the work of the Yale School generally) has much to offer. They would offer more, however, if there constructs at more surface levels (i.e., semantic and syntactic levels) were more flexible, allowing for variations across interpreters.
foster or hinder the consensual bases for political action?

One could approach these questions through intensive depth-psychological debriefings of interpreters who had undertaken the laborious coding task. But this approach would be enormously time-consuming. Alternatively, if the process of interpretation could be simulated, one could more readily develop a body of theory about the interpretive effects of presuppositions. In the process, techniques would be developed that would support the automated construction of linguistic structures from which plot units might be detected automatically, thus serving the first sense of hermeneutic in the Alker et al. study.

2.5.3 The Scope of Computational Hermeneutics

Although anti-hermeneutical models have been the rule rather than the exception in AI natural language research, the hermeneutical approach has recently gained momentum. Terry Winograd, a pioneer in AI natural language research, has been a prominent and vocal critique of the pre-programmed approach to interpretation. Expanding upon the Dreyfus brothers' incessant phenomenological critique of AI [Dreyfus and Dreyfus, 1984], Winograd [1980] introduced the AI community to hermeneutics and its implications for serious natural language implementations.

Winograd and Flores [1986] present the case against the application of pre-programmed expert systems as decision aids for human organizations. They argue that such applications cannot enhance the rationality of decision. Expert systems ordinarily lack the background
knowledge necessary to properly contextualize the decision situation. Unlike humans, they exhibit no "thrownness" — they do not interpret with the discretion and interpretive flexibility of a human decision-maker embedded in the social situations of the organization and the wider societies in which the organization itself is embedded. Also, they grant objective existence to the problems they are assigned to solve, remaining blind to the interpretive commitments made in the very formulation of those problems.

Winograd and Flores present insightful and well-grounded criticisms of existing AI applications to organizational decision-making. Their work is best seen as a necessary corrective for the overly optimistic appraisals for such applications that have recently appeared in the popular press.

AI is often anticipated by its boosters as some general panacea. It can indeed be helpful in making sense of the complexity that often confronts decisionmakers. It may also assist us in making decisions under uncertain conditions. However, the view of AI as a set of techniques for automating the frail, error-prone human component of decision is grossly mistaken. In emphasizing the incapacities of AI systems, such as their inability to respond to unforeseeable conditions, Winograd and Flores counter this naively optimistic perspective.

But we can realistically view AI as a set of techniques to assist decision makers, who might apply the techniques in order to make more responsible decisions. The Winograd and Flores critique of AI applies equally well to any modeling technique. To some extent because they
are not the phenomena they model, all models are woefully inadequate.

No model stands alone. Each requires, as a precondition of its usefulness, a human interpreter who construct the model and interpret it. Models provide model-interpreters epistemic access to the phenomenon being modeled. AI provides no additional, magical ingredient that frees humans from the responsibility for decision. Computational hermeneutics, then, holds promise for the development of systems to assist our interpretive understanding, not in replacing it with a new, pre-programmed authority.

A fully-blown, computational hermeneutic environment for the analysis of unrestricted textual content remains many years, perhaps decades, off. Non-trivial linguistic and computational issues must first be resolved. Nevertheless, research payoffs will accumulate as the technology develops toward that ultimate goal. As this work proceeds, it will afford ample opportunities for interim applications in areas of significant interest to social scientists. Examples of potential interim applications are: 19

1. **Historical Analysis** -- A hermeneutic computational environment would be able to base interpretations of historical narratives from a variety of interpretive assumptions. Differences between the deduced action consequences and precedent-based reasonings of different interpreters [Alker, Bennett and Mefford, 1980] could then be contrasted.

2. **Theory-Based Reasoning** -- In such an environment, one might represent a theory and apply it to individual cases. A propositional inventory of a social theory would first be

---

19 These applications and others (such as theory formation and social network analysis) are discussed in greater detail in [Duffy and Mallery, 1986], reproduced in Appendix I.
input in English. Textual accounts of particular cases are
next input, and the system then answers questions about the
cases. These answers constitute theory-based interpretations
of the story, which may be compared.

Analyses could also estimate the scope, internal structure,
logical consistency, and completeness of particular theories.
Propositions may also be added to or deleted from a
particular theory to see what differences these make in
interpretations of and inferences about the narrative.

3. **Intentional Game-Theoretic Analysis** — Game models are often
constructed as heuristics for understanding conflicts and
negotiations. The purely mathematical nature of most such
models, however, obscures the important intentional aspects
of players' game play, thereby affecting negatively the
validity of game-theoretic results. An hermeneutic
computational environment could incorporate players' responses to protocols designed to elicit their intentional
attitudes and attributions during the conduct of the game.
It might also recognize phases of play, and thus, classify
types of play.  

4. **Political Belief Analysis** — Studies of political belief have
thus far been either large-sample correlational summaries of
a narrow range of issue positions [e.g., Converse, 1964] or
non-rigorous interviews with individual respondents [e.g.,
Lane, 1962]. A computational hermeneutic environment would
allow the rigorous and replicable examination of the
political beliefs and inferences of individuals. One might,
for instance, test hypotheses concerning the addition or
deletion of beliefs.

As more computer scientists, following the lead of Winograd,
Flores, and others, become more attuned to the necessarily hermeneutic
character of interpretation, and as more hermeneutically trained social
scientists develop the computational sophistication to develop such
systems, these interim applications should begin to bear fruit.
Ultimately, as work progresses and as the relevant computational and
linguistic issues are ironed out, resolution of the validity problem of

---

20 Preliminary work in this area is discussed in [Alker, Duffy, and
Mallery, 1985; Hurwitz, Mallery, Alker, Duffy, 1986].
content analysis should emerge.

Alternative paths toward resolution of the validity problem may of course exist. The computational hermeneutic proposal presented here is just one proposal out of an infinity of possible proposals. I believe it to be the right approach, but perhaps it will stimulate discussions leading to a better approach.

2.6 Conclusions

When Pool summarized the Allerton House conference, resolution of the difficulties of computer text-understanding appeared to be just around the corner. Only after years of often frustrating efforts did more realistic appraisals of the enormity of the difficulties begin to appear. Our new-found appreciation of the obstacles need not lead us to despair, throw up our hands, and declare the problems irresolvable. While our knowledge of computer text-understanding techniques accumulates far more slowly than we might like, it has progressed and it continues to progress. The need for a hermeneutical approach to computerized text understanding only makes the problems more difficult, not impossible.

The issues that stymied quantitative content analysis do not necessarily disappear magically when we alter our focus from quantitative and statistical techniques to qualitative and linguistic ones. To the contrary, these issues remain at the foreground. If AI is to meet its promise as a set of tools for helping us to make better inferences about the world around us [Chan and Sylvan, 1984: 9], AI tools for testing hypotheses must be devised.
Alker [1975] locates AI within the universe of descriptive techniques. Computational hermeneutics, at least as I have characterized it here, can make possible tests of hypotheses by varying the interpretive assumptions behind readings of particular texts. How might the implications of an international event chronology differ for American and Soviet policymakers? How might the range of conceivable action possibilities expand or contract for a particular actor if such-and-such a belief were modified, deleted, or added? These are the sorts of questions that might be addressed.

By making hypothesis testing possible using such counterfactual techniques, the hermeneutical approach can supplement our methodological armamentorium with a set of AI-based inferential techniques. But, in order to serve as an environment in which hypotheses must be tested, the validity problem must first be resolved. Computational hermeneutics points the way toward such a resolution.

Because the interpretive backdrop against which texts are interpreted may be manipulated systematically, the computational hermeneutical approach advocated here would construct modeling environments capable of supporting hypothesis-testing. Hypotheses, on this view, concern the substantive effect on interpretation when a change is made in the backdrop of the interpretation -- whenever a particular proposition or set of propositions is manipulated systematically, whether added, removed, or otherwise altered.

The goal of testing, in a replicable fashion, hypotheses concerning differing interpretations across individuals and across social groups serves, it should be emphasized, two over-arching
purposes. First, its attainment would go a fair distance in implementing Moon's [1975] suggested synthesis of competing perspectives in political science — (a) the perspective of naturalists who seek general covering laws of social and political phenomena, and (b) the perspective of interpretivists focus on the cross-cultural differences that make political and social affairs more or less unique across disparate social groups and limit our ability to discover universal generalities. Second, because such a tool would foster the development of interpretively sensitive empirically-based social and political theories with broad applicability, they could serve to enhance the rationality of our practical political actions [cf. Alker, 1981].

Unlike previous computational approaches to content analysis, then, computational hermeneutics promises ultimately to be a tool for social-scientific theory construction that can surpass the limited applicability of methods borrowed from the natural sciences. And this, after all, is what Lasswell had originally sought.
Chapter 3

The Relatus Parser

3.1 Introduction

Upon deciding to investigate the possibilities and limits of the computational examination of text contents, a problem presents itself immediately: where to begin? Given that the research aims to simulate an aspect of human understanding, implementational efforts should be grounded in some coherent and defensive theory of human cognitive performance and development.

As will become abundantly clear in Chapter 4, I disbelieve approaches that propose a broad innate endowment of highly specific cognitive faculties. I instead believe an alternative which proposes that our innate cognitive heritage is rather more narrow, consisting of structures that support (a) a primitive degree of spatio-temporal differentiation, (b) an infantile ability to relate empirical objects to one another, and (c) the ability to learn from experience, including an ability to learn how to learn better. On this view, the various cognitive faculties that appear to be domain-specific competences, emerge developmentally from a more restricted innate endowment.

Discussion of this very controversial issue must be allayed in order first to discuss the choices actually adopted. We will assume for the moment that the developmental approach is substantially correct.

Now, how does one go about simulating human text-interpretation performance? One clearly requires some system for representing and interpreting knowledge. The frame-based systems that have been
developed from the seminal work of Quillian [1968] and Minsky [1974] seem particularly well-suited. They can perform the tasks necessary for effective knowledge representation and interpretation.

1. They relate symbolic entities to one another;

2. They implement the subsumption of particular terms by more general concepts in type hierarchies;

3. They implement the inheritance of properties from more general to more specific terms in those hierarchies;

4. They support the invocation of procedures when particular, prespecified configurations of relations are input; and

5. They support analogical pattern matching of relational configurations, thereby facilitating learning from experience.

Performance on these tasks (and others) varies across frame-system implementations, of course, but the important point is that they can and do perform them.\textsuperscript{21} Frame languages of this variety implement systems for the representation of knowledge that are consonant with epistemological approaches in the Western mainstream since Aristotle.

In Chapter 5, I defend a contemporary formulation of the classical approach against an important recent objection, and I show how it supports the formation of hypotheses in Chapter 6. Taken together, these chapters defend the proposition that the classical approach can support experiential learning. Of course, this proposition is defended much more cogently by Winston [1981], who produced in a frame-based environment a program that learns from examples. Chapter 6 describes

\textsuperscript{21}Winston and Horn [1984: 311–320] provide step-by-step instructions for constructing a version of Roberts' and Goldstein's [1977] FRL (Frame Representation Language) that can readily be made to incorporate these abilities for relatively small amounts of knowledge.
how creative learning might occur in such an environment through the formation of novel hypotheses.

In this chapter, however, I wish to discuss the Relatus Parser. On the assumption that we acquire the bulk of our syntactic knowledge through experience, embedding syntactic processing inside a general knowledge representation system would appear to make eminent sense. Because natural languages are constantly evolving, mutating, open systems, any natural language understander will need constantly to revise its beliefs about the appropriateness of particular expressions in particular contexts. It must continually form and test hypotheses about the language, reformulating, developing, and elaborating theoretical constructs\(^{22}\) in order both to improve its capabilities and to keep pace with changes in language usage.

At first blush, then, the most straightforward syntactic strategy would be to implement a parser (and generator) within a knowledge representation system. However appropriate this plan might seem, it was unfortunately not feasible for us. Available frame-based knowledge representation systems did not meet the technical requirements for representing syntactic structures and matching those representations in order to detect the similarities and differences between them.

One difficulty concerned paging. In order to infer parse rules,

\(^{22}\) Provisionally, we may consider these constructs to be representable as configurations of causal and intentional relationships among general terms (i.e., "noun phrase," "adverb," "anaphor," "amphiboly," "connotation," and the like) and contextual particulars (i.e., "scientific discourse situation," "political manifesto," "newspaper editorial," and the like.)
it would be necessary to represent syntactic structures at various transitional stages:

1. The words of a sentence and their order must be represented, along with the possible category (part of speech) assignments of each word.

2. The actual part of speech assignments must then be represented, so that it would be possible to infer rules for settling on the appropriate categories given the constraints of word order.

3. Each word in the sentence must then be assigned to a particular node representing a noun phrase, verb phrase, etc.

4. Each phrase must then be assigned to a node representing a clause.

5. Each clause requiring syntactic transformation into a canonical deep-structure form must then be linked to a node representing its transform, and its subsidiary phrasal nodes must be linked to their counterparts under the transform node.

6. Each constituent in each deep-structure phrasal node must then be linked to its referent in a semantic representation of the grammatical and logical relations of the sentence.

It readily became apparent that, even for the simplest of sentences, hundreds of nodes would be needed. Since it would be necessary to represent many such sentences in this way in order to infer rules for parse transitions, a representation incorporating thousands upon thousands of nodes would be needed. However, Mallory's experiments indicated that, with the maximum main memory then available (2.25 megabytes), paging performance began to deteriorate noticeably at about 1,000 nodes. At about 5,000 nodes, the machine spent almost all its efforts transporting data to and from disk.

Another technical problem concerned the time complexity of matching configurations of nodes. The number of nodes in the knowledge base increased linearly, the time consumed in matching increased
exponentially. Matching performance was slowed even further due to the paging problem.

These limitations prohibited any parser implementation inside available knowledge representation systems. Further, since any non-trivial political modeling application would also require the storage of and matches of large amounts of knowledge, it became evident that existing frame-based systems would not support political science applications. We would have to implement our own knowledge representation system so that memory could be managed and efficient matching could be implemented. The knowledge representation system, Gnosocere, is described very briefly in Appendix I and in more detail in [Mallery, 1987]. Further developments and greater detail will undoubtedly appear in Mallery's forthcoming doctoral dissertation.

In order to debug Gnosocere in its development phases, some means of interning data was needed. Since text representation was seen as the primary purpose of Gnosocere, it made sense to implement a parser that would analyze sentences, preparing their contents for mapping into the knowledge base. Section 3.2 discusses the implemented design of that parser, which converts sentence strings into graphical representations of sentence structures. Mallery subsequently implemented a constraint poster that posts reference constraints on nodes in the parse graph output and then exploits those constraints when mapping sentential contents into Gnosocere (see Appendix I and [Duffy and Mallery, 1984]. To further our long-term goal of learning parsing rules inside Gnosocere, the parser that was constructed can also map into Gnosocere descriptions of the parser's state at various key transition points.
Following the discussion of the parser design in section 3.2, section 3.3 describes an important component of the parser, a categorial disambiguator, which determines the categories (parts of speech) of particular words based upon the constraints imposed by other words in the sentence. Section 3.4 presents a number of parse graphs produced by the parser. Finally, section 3.5 discusses issues and problems relevant to further parser development.

3.2 The Implemented Design of the Relatus Parser

Parsers are notoriously complex beasts. Generally, as the range of acceptable inputs is extended, parsers become exceedingly complex and difficult to understand and extend. To counter this problem, and also to enable its principled integration with a semantic component, message-passing or object-oriented programming techniques were employed in implementing the Relatus Parser. This section describes the advantages of this approach in parser design.

The Relatus Parser converts English sentences to parse trees, the non-terminal nodes of which are message-passing agents. Three general advantages flow from this design choice:

1. **Generic operations** greatly enhance the clarity of coded operations, including operations for phrase structure creation, anaphora resolution, and transformation.

2. **Deterministic reference** is made possible. Each parser agent reasons about its own structure and its logical and grammatical relations to other agents. This facilitates reference in a semantic representation which minimizes expensive backtracking [Duffy and Mallery, 1984].

3. The **learnability** of procedurally encoded parse rules is enhanced. Each agent can be made a specialist in mapping its syntactic relationships into a semantic network model of syntactic structures designed for reasoning and learning.

These advantages are discussed below.
3.2.1 Generic Operations

Messages are generic operations. An agent receiving a generic message performs an operation specific to its type. This eliminates much type-checking, resulting in a modular distribution of knowledge without "hairy control structures" [Hewitt, 1976]. The system is thus much more easily understandable and more readily extensible.

Because message-passing was early adopted as a choice for the design of the Relatus Parser, the implementation has proceeded far faster than would otherwise have been possible. Its principled integration with a semantic network has also been expedited. Before discussing these advantages, the architecture should be presented in more detail.

3.2.1.1 A Typology of Parse Agents

The Relatus Parser may not constitute the first object-oriented parser. Phillips [1983] presented his design as the Relatus Parser was being implemented. Phillips, however, defines only one type of agent, the "constituent", while the Relatus Parser defines nine distinct, instantiable agents. It also defines many more non-instantiable agents, the state and functionality of which the instantiable agents inherit. As a result, the Relatus Parser exploits the power of message-passing techniques to a much greater degree than does Phillips' parser.
3.2.1.1.1 Instantiable Agents

The implementation deploys two clusters of instantiable agent types — three clausal agents and six phrasal agents. Each agent includes a slot which points to its parent in the parse tree.

All instantiated clausal agents encompass phrasal agents within their scope.

1. **Sentence** agents represent matrix clauses.
2. **Kern** agents represent embedded, clauses.
3. **Transform** agents represent syntactic transformations of a sentence, kern, infinitival, or another transform.

An instantiation of a phrasal agent represents a sub-clausal phrase. Some may encompass both clausal agents and/or phrasal agents within its scope.

1. **Nominal** agents represent noun phrases. They may embed kern agents on their noun slots, and they may embed kern (dependent clause), nominal (bound prepositional phrases) or adverbial (adverbs which modify adjectives) agents in the list on their adjective slots. Junctional agents may appear on a nominal's noun slot, prepositional slot, or adjective list, whenever a noun, preposition, or adjective phrase, respectively, is conjoined or disjoined.

2. **Verbal** agents represent verb phrases, including the verb, verb root, tense, mood, aspect, and any negation operators.

3. **Adverbial** agents represent adverb phrases. They embed any adverbials which modify their adverb. Kerns that adverbially modify a verb (independent clauses) also appear on the adverb slot of an adverbial.

4. **Junctive** agents represent conjunctions and disjunctions of any agent type. They can conjoin or disjoin clausal agents, phrasal agents, or terminal symbols. They have a junctive slot, which indicates the type of juncture, either AND, OR, BUT, NOR, XOR, or XNOR.
5. **Comparative** agents represent comparisons (e.g., "more than" or "less than" relations). The creation of an comparative places the clausal parent immediately scoping it on a queue for later transformation to a syntactically canonical form.

6. **Complement** agents represent words or phrases that signal the boundaries between clauses. Examples are "because," "and," "that," so that," "as soon as," etc.

### 3.2.1.1.2 Non-Instantiable Agents

Each instantiable agent inherits slots and operations defined on certain non-instantiable agents. In this way, slots and operations shared across agent types need only be defined once — on a non-instantiable agent. Figure 3-1 displays the skeletal agent hierarchy for clausal agents while Figure 3-2 displays the skeletal hierarchy for phrasal agents.

![Diagram](image)

**Figure 3-1: The Skeletal Hierarchy of Clausal Agents**
Figure 3-2: The Skeletal Hierarchy of Phrasal Agents

All agents in Relatus inherit the "abstract" Basic Agent, which provides operations common to all. These include operations for agent self-description, documentation, and for dumping to binary or textual files.

Not shown are other "abstract" agents that provide specialized operations to the instantiated agents. By defining them on abstract agents, specialized operations are grouped together in modules so that the implementor may find them easily. This convention enhances code clarity and speeds debugging significantly.

Additionally, this modularity allows developers to define protocols on abstract agents. Whenever a new protocol is implemented, the abstract agent can be "mixed out" of the instantiable agent's definition, while a new abstract agent, defining the new protocol, can be "mixed in". Such swapping of modules may be restricted to a subset of the instantiable agents that inherit the abstract agent. Thus, some agents can be made to inherit the new protocol, while others inherit
the old one, with no conditionalization of the code which invokes the operations.

Further levels of genericity and code organization flow from the use of "macro" and "daemon" operations. For example, in response to a particular message, an operation defined on a non-instantiable agent may be "advised" by a "macro" or "daemon" operation defined on an instantiable agent to perform certain tasks prior to and/or after the invocation of the main operation. This technique has proven useful in avoiding code duplication whenever agent responses to messages are similar but not exactly equivalent.

3.2.1.2 Overview of the Parse

The Relatus Parser was implemented to provide textual data for developing, testing, and debugging a semantic network which implements a theory of semantic perception [Mallery and Duffy, 1987]. The parser takes the following actions:

1. The parser's resource manager allocates a sentence agent.

2. The sentence agent constructs a list of all possible categories (parts of speech) for each word, based on information in a lexicon.

3. The sentence agent inserts missing words from short ellipses. For instance, "all the men" becomes "all of the men". Deleted complements are also inserted here. For example "Oliver North knew he was in trouble becomes "Oliver North knew that he was in trouble."

4. The sentence agent disambiguates the possible categories using a constraint-propagation technique, described in section 3.3.

5. The sentence agent clause supervises a phrase-structure parse within its scope, and saves a list of phrasal agents.
6. Clausal boundaries are detected by searching in the list of phrasal agents for complement agents. The complement agent then supervises the embedding of a kern inside the sentence, which keeps the agents that precede the complement. Succeeding agents are handed to the kern, which then recursively detects further clausal boundaries. The nature of the embedding in the higher level, or matrix, clause is determined by the complementizing work or words encashed by the complement agent and by the types of the phrasal agents that precede the complement agent. When necessary, on agent in the matrix may be copied and lowered into the kern. The agent and its copy are marked as coreferents so that the constraint poster and reference interpreter will know that they will both refer to a single mode in the knowledge representation.

7. A phase of intrasentential coreference resolution, discussed below, occurs.

8. Any clause requiring transformation is transformed. The phrase structure in the clause is copied and converted to a syntactically canonical deep-structure form.

9. Reference constraints are posted on the non-terminal nodes of the deep-structure parse agents. These constraints are then collected and passed to a reference interpreter that merges the sentence contents into the knowledge base [Mallery, 1987].

3.2.2 Advantages of the Object Orientation

Aside from advantages gained from clean control structures, code organization, and code understandability, message-passing simplifies procedures particular to the parser and to natural language parsing in general.

3.2.2.1 Simplifying Deterministic Reference

Arguing from efficiency grounds, Duffy and Mallery [1984] present a Deterministic Reference Principle, which advises implementors to "prefer a natural-language understanding architecture which maximizes the monotonicity of reference". They argue that explicit syntactic
deep structures help satisfy this principle. The object-orientation of the Relatus parser also helps to satisfy it.

Once the parse terminates, a constraint-posting phase begins. This process, detailed in [Mallery, 1987], implements the linguistic level of logical form [Chomsky, 1981: 324]. Briefly, a "post-constraints" message is sent to all parser agents in a depth-first walk of the parse tree. Each type of instantiated agent responds differentially to this generic message, reasoning about its local contents and posting on itself constraints that reflect its logical (including quantificational) and grammatical relationships to its neighbors and constituents in the parse tree. The constraints are then composed into a reference description of the entire sentence, which then maps the contents of the sentence into a semantic network. Sentence constituents that refer to nodes already represented in the semantic representation are merged into those nodes, thereby eliminating much later backtracking to merge coreferring nodes.

Constraint-posting decisions are made locally. Each parser agent is a specialist in posting constraints within its own domain. This greatly simplifies the process of constructing a principled specification of the sentence suitable for mapping sentential contents into a semantic network.

3.2.2.2 Simplifying Coreference Resolution

While intersentential coreference is resolved in Gnoscere, intrasentential coreference is handled exclusively by the parser. The parser resolves two sorts of coreference: those due to anaphor resolution and those due to "lowering" operations. In either, when two
nominal agents corefer, each nominal records the other on its
coreference slot. Whenever more than two nominals corefer, a cycle of
coreference is established. An operation is defined on nominals which
returns a list of all its coreferents. The constraint poster exploits
this information, ensuring that the coreferent nominals will also
corefer in the semantic network. This also helps to satisfy the
Referential Determinism Principle.

The representation of parse-tree nodes as message-passing agents
greatly simplifies pronominal anaphor resolution. When a pronominal
nominal agent is created, it is pushed onto a queue for resolution
after the phrase-structure parse terminates and before any
transformations. Depending upon the class of anaphor and its position
within the nearest governing clausal agent, a particular search
strategy is selected. Search proceeds outward from the anaphoric
nominal. Thus, only the range of the parse tree is searched in which
coreferent nodes logically can be embedded. Agents that cannot resolve
the anaphor invoke no-op operations when they receive the message. If
they embed syntactic structures, however, they pass the message along.

When clausal constituents are created, noun phrases must sometimes
be "lowered" to the embedded clause from the parent clause. For
example, the sentence "Elmer wants to fly a kite" becomes

\[(\text{Elmer want (Elmer fly kite)})\].

When this occurs, the nominal agent (Elmer) "lowered" is made to
corefer with its coreferent in the higher clause. Like anaphor
resolution, this reasoning about coreference occurs locally.

3.2.2.3 Simplifying Transformations

Each clausal agent, including the transform, maintains a slot for a transform. A transform is a canonical representation of the grammatical relations in each clause. Only transformations that move constituents across phrasal boundaries are represented in this way. For efficiency reasons, local transformations are performed during the earlier, lexical insertion phase. When the parser creates a phrase structure requiring transformation, the nearest governing clausal agent is placed on a transformation queue.

Transforms copy the top-level phrasal agents of the clause to be transformed and maintain a map between the transform's phrasal agents and the corresponding agent in the clause. When the constraint-poster sees that a clause has been transformed, it analyzes the transform instead, posting constraints which represent the canonicalized grammatical relations. Since quantifier scopes must be analyzed in surface structure [Jackendoff, 1972], phrasal agents inside transforms simply send a message to their counterparts in the surface clause asking for its quantification constraints. The surface phrase reasons about its quantification locally, returning the answer to its deep-structure counterpart, which posts it appropriately.

I lied about the trees. The parser sometimes produces a cyclic graph. It does so whenever a transform appears at the matrix level or at some intermediate level and other clauses are embedded more deeply.

Transformation operations copy only the top-level phrasal agents
of a clause. Because the copied agents receive the slot bindings of these constituents, the phrasal agents of a clause and those of its transformation share constituents, preserving eq-ness. Thus, when constraint posting proceeds downward from a transform in deep structure, it automatically returns to surface structure in the transform's constituency.

Figure 3-3 displays the parse graph for "Elmer wanted to give Agnes the dog that he bought". In this sentence, the constraint poster's path branches from the surface (kern-1) to its deep-structure dative transformation and returns to surface structure in kern-2, an adjectival constituent of both nominal-5 in kern-1 and nominal-9 in the transformation.

So, although the parser produces a cyclic graph of clausal and phrasal agents, the constraint poster actually walks down a tree of syntactically canonical clauses interwoven in that graph. When no transformation is present, it proceeds downward through the phrasal constituents of the surface clause. When there is a transformation, the constraint poster simply shifts over to the transform agent, proceeds downward through its phrasal constituents, emerging in surface structure at the next lower clausal level.

---

23 The slot binding for the parent of the copied agent is changed, however, to point to the transform rather than to the surface clausal agent. Additionally, in passive transformations, the preposition slot for the "by" prepositional phrase is bound to NIL.

24 "Eq-ness" means sharing a memory location. In Lisp, if two things are eq, they are physically one and the same thing.
Figure 3-3: Parse Derivation for a Complex Sentence

a. Surface Structure

b. Deep Structure
There is no need, after posting constraints at deep structure, to return to the surface clause and search for additional constituents at lower levels in the derivation. Constraint posting is thereby simplified. While it is certainly possible to use this cyclic technique in a list-structure syntactic representation, an object orientation makes the cycle explicit and therefore easier to debug. Since agents have unique printed representations (unlike lists) one need only eyeball the parse output to tell whether a constituent is the same at both surface and deep structure or has been copied. For any graph-walking procedure, such as constraint posting, no elaborate bookkeeping need be maintained concerning the current and previous locations of the traversal. Nodes in the parse graphs simply pass messages to each other.

One might object that the parser's creation of deep structures proliferates the size of the derivation, resulting in inefficient space consumption. Although in [Duffy and Mallery, 1984] and [Mallery and Duffy, 1987] arguments are presented for the necessity of deep structures, the object-orientation of the parser already makes this space-consumption objection moot. All parser agents are created at login time, and encached inside a resource manager. When a parser agent is needed, the resource manager simply pops one off a list. When the parse graph is no longer needed, the resource manager pushes the agents back onto the list. The resource manager thereby acts as the parser's own, local, garbage collector, minimizing its total space consumption. When parse derivations are represented in list structure, resourcing is far less feasible.
3.2.2.4 Simplifying Copular Quantification

In copular clauses, noun phrases quantify differently than other clauses, because copular verbs identify and categorize [Mallery, 1985]. The message-passing control structure facilitates localized reasoning about copular quantification.

In copular clauses, the correct quantification of each noun phrase depends not only upon its own number and determiners, but also upon the numbers and determiners of other noun phrases in the clause. The quantification of each is mutually determined. Because noun phrases are message-passing agents, each can query any of the others about its number and determiners simply by sending a message. Operations inherited by nominal agents perform this mutual determination of quantification [Mallery, 1985], implementing and extending the analysis in [Jackendoff, 1983].

3.2.2.5 Data Creation for Learning Declarative Parse Rules

Syntactic theories that posit mechanisms outside the semantic component are vestigial remnants of the "autonomy of syntax" thesis. Parsers constructed without consideration of their principled integration with a cognitively plausible semantic component stand little chance of being cognitively plausible themselves [Duffy and Mallery, 1984; Mallery and Duffy, 1987]. But this does not suffice. Because syntactic knowledge must be learnable, a plausible model of

---

25See also Chapter 4.
syntactic performance should be embedded within a cognitively plausible semantic mechanism designed for reasoning and learning.

Syntactic components constructed independent from and outside of the semantic component constitute an unnecessary level of machinery. Marcus [1980: 226] argued that the need for semantic operations, such as inquiries to resolve prepositional binding ambiguities, indicates that the two should not be kept artificially separated.

An even more compelling case for an integrated syntactic theory is based on the learnability criterion. Because most syntactic rules must be acquired, a cognitively plausible parser must be embedded within a learning mechanism. Since learning depends critically upon sense and reference, the learning mechanism must be embedded in the semantic representation. This topic is treated in great detail in Chapter 4.

The Relatus Parser could not be implemented inside a semantic representation precisely because it was to be used to generate data with which to develop, test, and debug this component. A conscious decision was made to implement a procedurally encoded parser which can be mapped into a semantic network representation.

The object-orientation of the parser makes this bootstrap plan feasible. At various critical moments during the course of a derivation, parser agents can reference themselves in a semantic network. They do so by sending themselves (and each other) generic messages that construct reference specifications describing their relationships to words, categories, and other parser agents. Because

\[26\] An existing parser or representation system might have been used. But, in general, because they do not satisfy the Referential Determinism Principle, available existing systems would not suffice.
generic messages perform this syntactic self-reference, they need never reason about where they are in the derivation. They need only pass messages to themselves and to their neighbors.

Once constructed, the reference descriptions are mapped into a semantic representation. Since constraint-interpreting reference merges items with coreferring nodes in the semantic representation [Mallery, 1987], a consistent model of the transitions of the syntactic derivation is constructed incrementally. All syntactic transitions are modeled, including categorial disambiguations, lexical insertions, clausal and phrasal scopings, and transformations.

Once the inferential machinery planned for Mallery's Gnosore is fully installed we hope to experiment with representations of the transition states of the parser, learning parse rules from these transitions. By inhibiting certain rules and enabling others in particular discourse contexts, we hope to be able to process texts written in various dialects and ideolects. We also hope to run these rules (conceptually) in reverse, so that from the representation of the transition states of a parser, we will be able to generate an English-language generator. As well, we hope eventually to be able to state in English the transition rules for other languages, in order to facilitate the construction of parsers and generators in those languages.

3.2.3 Design Summary

Message-passing techniques have helped speed the development of the Relatus Parser in many ways. Generic messages have simplified the control structure. Abstract agents have helped to organize the code
for easy retrieval. Agent-centered representations have facilitated the development of elegant strategies for the consistent and principled integration of the parser with semantic representations of both textual contents and syntactic structures.

Message-passing is highly recommended both as a processing strategy and as a representational strategy for serious parser implementations.

3.3 Categorial Disambiguation

Ambiguities pervade natural language. Many strategies exist for resolving some of the varieties of ambiguity, including phrasal and clausal attachment ambiguities, anaphor ambiguities, referential ambiguities, and sense ambiguities. Categorial ambiguities arise when the lexical entry for a token (word or morpheme) indicates that the token may be given alternative category (part of speech) assignments, depending upon the context of usage.

Few strategies exist for resolving categorial ambiguities, the simplest lexical ambiguity of all. Most existing approaches resolve categorial ambiguities either (a) by following all categorial parsing paths until a grammatical path terminates, or (b) as part of the process of resolving sense ambiguities. Both approaches are computationally expensive. This paper presents the alternative approach that has been implemented in the Relatus Parser.

The categorial disambiguator described here constitutes a simple method for resolving categorial ambiguities before phrase structures
are created. As a result, only one parsing path need be followed and any later sense disambiguation is greatly simplified.

3.3.1 An Overview of Categorial Disambiguation

The categorial disambiguator assigns the appropriate category (part of speech) to a word whenever more than one such assignment is possible. For example, in sentence (1) below, the words "doctor", "might", "cure", and "patient" are each categorially ambiguous. Both "doctor" and "cure" could be a noun or a verb. "Might" could be a noun or a verb auxiliary. And "patient" could be an adjective or a noun. In (1), only the two instances of the definite article "the" are not categorially ambiguous. Yet, as we shall see, these suffice to disambiguate the category assignments for the entire sentence.

(1) The doctor might cure the patient.

When considering a design for the categorial disambiguator, an immediate inspiration was Waltz' [1975] constraint-propagation approach for detecting legal junctions in line-drawings. Applying this approach to the detection of legal category combinations in English sentences proved to be quite straightforward.

Prioritized pattern action rules (CONDs) are specified for each known categorial ambiguity. Approximately 40 such rules are currently in place. Each rule is passed lists of the words and categories preceding and succeeding a categorial ambiguity. Other categorial ambiguities are represented as lists embedded within these preceding and succeeding category lists. Unambiguous categories are represented as symbols in those lists. Additionally, each rule knows the current
word and maintains a list of its possible category assignments, which had previously been extracted from a lexicon.

When a disambiguation rule succeeds, it propagates its resolution as a constraint for disambiguating neighboring ambiguities. When a rule fails to resolve the ambiguity, the ambiguous alternatives remain in the category lists. When subsequent disambiguations propagate additional constraint, the disambiguation rule for this ambiguity is re-evaluated.

Ordinarily, disambiguation rules need examine only the categories of its sentential neighbors. These usually provide sufficient constraint to select one correct interpretation. Occasionally, however, the words in a sentence must be consulted in addition to their possible category assignments. For example, deciding whether a particular word is an adjective sometimes depends upon knowing whether a preceding verb accepts predicate adjective arguments. For maximal flexibility, disambiguation rules are free to query the lexicon about syntactic and semantic properties (subcategories) of particular words.

3.3.2 An Example of Categorial Disambiguation

The disambiguator resolves sentence (1) in the following steps:

1. "The" is unambiguously a determiner.

2. "Doctor" cannot be a verb because it follows the determiner "the". It must therefore be a noun.

3. "Might" might attach to "doctor" genitively, just as "door" sometimes attaches to "car" to form "car door". Since we are working without information regarding sense, we cannot reject this possibility. Thus, "might" remains categorially ambiguous. It is either a noun or a verb auxiliary.

4. "Cure" as a verb is consistent with the interpretation of "might" as a verb auxiliary. If "might" were a noun, then
"cure" could not be a verb, since it disagrees in number with "might" as its subject. Of course "doctor might cure" could itself be a genitive formation, much the same as "car door handle". But if this were the case, then the next word, "the", would be anomalous. Two independent noun phrases rarely appear before a verb. "Cure" must therefore be a verb. Since "cure" is a verb, "might" must be a verb-auxiliary. Otherwise, the number of "cure" would be inconsistent with the number of either "doctor" or "might" as the clausal subject.

5. "The", again, is unambiguous.

6. "Patient" follows the determiner "the", so it could still be either a noun or an adjective. However, "patient" terminates the current clause, so the analysis of "patient" as a noun is preferred.

3.3.3 The Fallibility of Categorial Disambiguation

The careful reader will have noticed that categorial disambiguation is fallible. Two possible sources of error appear in sentence (1). First, "cure" might indeed be a noun, since two independent noun phrases sometimes do appear before a verb phrase, when a complement has been deleted, as in sentence (2).

(2) Men dogs bite scream.

In (1), the rule that disambiguates "cure" can check the current clause for other possible verbs. Since "doctor" has already been disambiguated as a noun, no other possible verbs remain in the clause. Thus, the interpretation of "cure" as a verb is preferred.

In (2), when the rule that disambiguates "dogs" (which might be a noun or a verb) examines the clausal environment, it notices that two verbs -- "bite" and "scream" are available in the current clause, warranting the nominal interpretation of "dogs".
Actually, this case is somewhat more complex, since both "bite" and "scream" are themselves categorically ambiguous. Each could be either a noun or a verb. The ambiguity is resolvable, however. Since "men" is unambiguously a noun in (2), and since the three remaining tokens are all doubly ambiguous, there are $2^3 = 8$ possible sequences of noun phrases (NP) and verb phrases (VP) in this sentence. Only one of these sequences -- NP NP VP VP with a deleted complement between the two NPs -- is grammatical.

Both of these examples involve secondary searches through the clause. These occur only for a very limited range of cases. For example, had the verb auxiliary in (1) been "would", instead of "might", such a search would not be needed. In (2) the presence of an auxiliary, a determiner, or a complement might obviate such a secondary search. On the basis of a worst-case analysis, one might consider this practice explosive, but since such secondary searches are rare and since most clauses are not extremely long, the disambiguation procedures terminate quickly in practice.

The second possible source of error involves sense ambiguities masquerading as categorial ambiguities. For example, the interpretation of "patient" as an adjective in (1) might certainly be sensible. Since it could imply that the doctor never treats the impatient, (1) is semantically ambiguous. The categorial disambiguator assumes no ambiguities of sense. This might be seen as a weakness in the approach. In a system that includes a full-fledged discourse component, however, the disambiguator could be made sensitive to discourse cues indicating an adjectival interpretation for such cases.
Perhaps the most important source of the fallibility of categorial disambiguation is the fact that no exhaustive set of disambiguation rules exists. Categorial disambiguation relies on an as yet unarticulated theory of category combination constraints. This being the case, the disambiguator has developed incrementally, through practice.

First, an initial set of the most intuitively obvious pattern-action clauses was constructed for the most common categorial ambiguities. Whenever sufficient constraint was unavailable using these rules, and whenever the disambiguator selected a category incorrectly, users were asked to select the appropriate category. These selections triggered a background mail process that reported the sentence, the particular ambiguity, and the user's selection to the implementor. This information proved indispensible in developing a large set of rules covering a broad range of ambiguous conditions.

3.3.4 A Categorial Disambiguation Rule

One simple categorial disambiguation rule decides whether a word (in this case only the word "to") is a preposition or the infinitival complement. It is presented schematically in Table 3-1.

Each condition evaluates sequentially until one succeeds. The correct answer (rule consequent) replaces the alternatives in the category lists. Conditions 1 and 2 handle the simplest cases, in which the following categories unambiguously indicate a verb phrase or a noun phrase. Condition 3 handles split infinitives. Conditions 4 and 5 handle cases in which the word is at the end of a sentence or clause. If the clause begins with a subset of WH complements (who, which, what,
where), "to" is considered to be a preposition (something up with which Winston Churchill would not put). Otherwise, "to" is made the infinitival complement (with an ellipsis). Conditions 6 and 7 handles cases in which the next element is also ambiguous, but is either a noun or a verb. If the next word would be a singular noun (e.g., "to store"), "to" must be a complement. If it would be a plural noun (e.g., "to stores"), "to" must be a preposition.

<table>
<thead>
<tr>
<th>Condition</th>
<th>Consequent</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. unambiguous noun phrase follows</td>
<td>preposition</td>
</tr>
<tr>
<td>2. unambiguous verb phrase follows</td>
<td>complement</td>
</tr>
<tr>
<td>3. adverb follows with following verb phrase</td>
<td>complement</td>
</tr>
<tr>
<td>4. WH clause termination point</td>
<td>preposition</td>
</tr>
<tr>
<td>5. non-WH clause termination point</td>
<td>complement</td>
</tr>
<tr>
<td>6. next word is either a singular noun or a verb</td>
<td>complement</td>
</tr>
<tr>
<td>7. next word is either a plural noun or a verb</td>
<td>preposition</td>
</tr>
<tr>
<td>8. disambiguation failure</td>
<td>alternatives</td>
</tr>
</tbody>
</table>

Table 3-1: Rule for disambiguating "to"

Condition 8 is the failure condition. The undisambiguated alternatives are returned. Rule failures always return the ambiguity. In such cases, succeeding disambiguations may propagate enough constraint so that a reapplication of the rule would successfully resolve the ambiguity. Only when no possible additional sources of constraint exist will an ambiguity be considered irresolvable. In such cases, the user is asked to resolve the ambiguity, and the user's action is recorded so that the rule might be extended.

3.3.5 Heuristics for Rule Construction

The disambiguation rules are designed to minimize the amount of computation needed for successful resolution. Rules for ambiguities
involving two alternatives attempt to rule in the correct category. Generally, for two alternative rules, conditions which are least computationally expensive to evaluate are evaluated first, while more expensive conditions are evaluated later, if necessary.

Rules for ambiguities involving more than two alternatives attempt to rule out, rather than rule in, particular alternatives. For efficiency, alternatives considered least likely or easiest to rule out are checked first. If an alternative can be ruled out, control passes to the rule that disambiguates the remaining alternatives, and so on, until the ambiguity is resolved. Ruled-out alternatives do not reappear on the list of alternatives for a particular word in a particular sentence. Instead, only the remaining alternatives appear for possible later disambiguation.

3.3.6 Assessment of the Current Model

Categorial ambiguities are currently resolved prior to the detection of phrasal boundaries. The speedy development of a disambiguator with a broad range of coverage motivated the choice to implement it as a module separate from other parser components.

There is no reason, in principle, why the disambiguator could not be more tightly integrated with the parser. To do this, however, a stack of previously parsed words and categories would need to be maintained. Information regarding their linear order might be lost in the phrase structures. Also, since decisions regarding the appropriateness of particular parse rules often depend upon knowledge of succeeding categories, an integrated implementation would require occasional resolutions of categorial ambiguities ahead in the sentence.
As is well-known from expert-systems research, changes in one rule can sometimes produce unexpected negative results from other rules that were not changed. The categorial disambiguator is by no means free of these problems. They are largely eliminated, however, by (a) avoiding side-effecting consequents in the rules even when a successful condition might warrant the immediate disambiguation of category elsewhere in the sentence, relying instead on constraint propagation, and (b) the installation of debugging tools that trace the evaluation of rules. As a result, the causes of failed disambiguations are simple to detect and thus also to remedy.

Variations on the current approach are certainly possible and might constitute interesting lines of research. For example, one might implement an expert-system variant, in which disambiguation rules are represented declaratively and perhaps weighted to indicate their relative values evidential impact upon a resolution. Such an approach would undoubtedly run slower than the prioritized procedural approach described here, but the model would also be more easily manipulable and applicable to other languages.

One might also attempt to remove the implementor from the debugging loop, implementing a variant which automatically learned new rules from disambiguation failures. The practical difficulty of this approach, however, involves the automatic detection of the reasons for such a failure. Errors might be due, for instance, to the specific subcategories of particular words. Without a full-blown theory of the role of subcategories in categorial disambiguation, it is difficult to
see how a program could be made to select the correct subcategory consistently.

3.3.7 Related Work

Very little research seems to have been conducted on the resolution of categorial ambiguity. This has been somewhat surprising, since the technique is quite straightforward and the results are most powerful.

3.3.7.1 Milne's Disambiguator

As an earlier version of this chapter was in press [Duffy, 1986], a report of Milne's [1986] categorial disambiguator appeared. Milne's disambiguator is far less powerful than the one described above, covering only a few of the possible ambiguities. But Milne was not interested in the breadth of its coverage. Instead, he wanted to show how the rules for categorial disambiguation\(^{27}\) can be integrated into Marcus' [1980] deterministic parsing model. Thus, Milne's disambiguation rules are also deterministic in the sense of Marcus. This, believes Milne, provides his categorial disambiguation procedures with cognitive plausibility.

As the discussion in Chapter 4 indicates however, cognitive plausibility cannot be granted to any stand-alone parser. It must be embedded within some cognitive processing model that can parse the world as well as a sentence. Ultimately, one would indeed like a model of categorial disambiguation embedded inside rules for syntactic

\(^{27}\)Milne does not use this term.
processing, but these rules should themselves be embedded within a structure capable of acquiring those rules.

3.3.7.2 Wilks' Preference Semantics

As part of his "preference semantics" approach, Wilks [1973] resolves categorial ambiguities by characterizing sentences as alternative sequences of semantic primitives and testing each for goodness of fit using a database of templates expressed in those primitives. To use one of Wilks' examples, sentence (3), in which the term "father" is categorially ambiguous, may be characterized as two alternative sequences of semantic primitives, (3a) and (3b).

(3) Small men sometimes father big sons.
(3a) KIND MAN HOW MAN KIND MAN
(3b) KIND MAN HOW CAUSE KIND MAN

The alternative interpretations are then reduced to a sequence of "bare templates" by stripping off the adjectival KIND and the adverbial HOW, resulting in MAN MAN MAN and MAN CAUSE MAN. These two sequences are then matched against a database of legitimate bare templates. Since only the second sequence appears in that database, (3b) is correctly chosen as the appropriate interpretation.

It is important to note that Wilks' technique was not designed to resolve categorial ambiguities. It does so as a by-product of its resolution of sense ambiguities. This does not mean that sense disambiguation, by this method or by any other (e.g., [Small, 1980]) obviates categorial disambiguation. Quite the contrary, it means that categorial disambiguation can be an inexpensive prelude to sense disambiguation. Had sentence (3) been categorically disambiguated
earlier, "father" would have been recognized as a verb, so the nominal usage of "father" would already have been ruled out. No reduction to primitives and no matching in a template database would have been needed.

3.3.7.3 Breadth-First Parsing

Another alternative method for resolving categorial ambiguities — pursuing all categorially possible parse paths in breadth-first fashion [Martin, Church, and Patil, 1981] — is exponential in the number of categorial ambiguities. The number of alternative parse paths which would be tried is the product of all possible category assignments. For sentence (1), for example, the number of alternative paths would be:

(1) The doctor might cure the patient.
Alternatives: 1 x 2 x 2 x 2 x 1 x 2 = 16

Worse, many categorial ambiguities involve three or four possible category assignments. For example, some words can be progressive verb forms ("You are winning"), gerundive adjectives ("the winning entry"), or gerundive nouns ("Winning is everything"). Extremely common words, like "in", take even more possible category assignments.

3.3.7.4 Phenomenologically-Plausible Parsing

Waltz and Pollack [1984] present a connectionist model in which categories are disambiguated concurrently with other lexical ambiguities. Unlike the serial processing case, prior resolution of categorial ambiguities in connectionist models would not result in any significant time savings. In fact, in a connectionist model,
sequential resolution of categorial ambiguities prior to the resolution of other ambiguities would consume more time. There is a trade-off, however. As in the serial case, earlier categorial disambiguation would constrain the range of possible sense interpretations. Instead of saving time, in the case of massively parallel processing, prior categorial disambiguation would conserve processors.

3.3.8 Summary

Categorial disambiguation is a computationally inexpensive means for reducing the ambiguity of a sentence. While categorial disambiguation does not resolve all the ambiguities that may appear in a sentence (e.g., anaphor ambiguities and ambiguities of prepositional attachment, clausal attachment, and sense), it does eliminate a source of ambiguity that pervades sentences. Moreover, the prior resolution of categorial ambiguities radically simplifies any procedures that resolve these other classes of ambiguity.

The implemented disambiguator remains in a development stage. No tests have yet been performed, using representative texts, to estimate its degree of coverage. Nevertheless, experience with moderately complex sentences indicates that the current set of rules is quite robust. As disambiguation failures are detected, the rulebase is extended, making it more robust. Most importantly, its efficiency over breadth-first approaches and the leverage it provides for other forms of disambiguation suffice to warrant use and extension of the technique.

As a final note, a large amount of syntactic knowledge is embedded within the categorial disambiguation rules. These rules are also a
potential source of considerable constraint for the resolution of ambiguous morpheme sequences output by a speech recognition program. By applying a categorial disambiguator to this output, syntactic constraints on possible morpheme sequences may be applied without the overhead involved in testing alternative parse-tree constructions.

3.4 Parse Output

This section examines and explains several of the parse graphs that the parser produces when handed a sentence. These graphs do not exercise the full capabilities of the parser, only some of the more important capabilities. The sentences that are parsed are meant only to present the parser's output. They may or may not bear any resemblance to reality.

Readers conversant with syntactic analysis are forewarned that they may well become impatient with this discussion, since it explains how the parser performs some basic movement operations without presuming technical knowledge of linguistic theory. The graphs are presented in Figures 3-4 to 3-13, which are collected near the end of this section.

Some terminology must first be clarified. There are several classes of clauses. The clausal agents, "sentence," "kern," and "transform" are described above. Transforms always represent the canonicalized (declarative) deep-structure version of a sentence or a kern. Kerns are always embedded clauses, and sentences are always the top-level clause. As used below, the term "matrix" refers to the clause in which a particular kern is embedded. A matrix clause may therefore be a sentence agent or a kern agent. If a matrix clause has
been transformed, then any kern embedded within it will have more than one matrix clause -- the surface matrix and any transformations of the surface matrix. In general, when referring to a kern's matrix which has been transformed, the surface matrix is implied unless otherwise stipulated. Also, references to "the matrix" without simultaneous reference to a kern denote the top-level sentence agent.

The parsed sentences described here are not meant to be representative of texts generally, and they are not drawn from any particular text. Instead, they were constructed to illustrate some of the features of the Relatus Parser.

Figure 3-4 presents the surface structure parse graph produced for the sentence "North is the man Meese said was giving the Contras financial assistance illegally." Notice that the lexical insertor placed "who" between "man" and "Meese" in order to make the sentence intelligible. The embedded kerns, KERN-1 and KERN-2, each embed within their matrices according to the functional rule each performs in its matrix. Thus, KERN-1 is embedded as an adjective of NOMINAL-2 in the matrix, while KERN-2 is the noun of NOMINAL-4 in its matrix -- KERN-1.

The dotted line between NOMINAL-2 and NOMINAL-5 indicates that the two nominal agents corefer. That is, the "man" on the noun slot of each nominal refers to the same man. Intrasentential coreference links of this sort are made whenever an intrasentential pronominal is resolved and, as in this case, whenever a noun phrase is "lowered" across clausal boundaries. This sentence is rather unusual in that the noun phrase "man" (NOMINAL-2) is lowered across two clausal boundaries.

In some sentences, more than two noun phrases corefer. In such
cases the nominals point to each other in cyclical fashion. When applied to a nominal agent, a COREFERENCE-CYCLE method takes an argument that specified whether the caller wants the surface structure or deep structure coreferents, or both. The constraint poster exploits this information so that when reference constraints are interpreted they will pick out the same "man" in the knowledge representation.

Notice in Figure 3-4 that the adverb slot of ADVERBIAL-1 inside KERN-2 encodes "illegal" as the adverb rather than "illegally." The code that creates adverbials canonicalizes adverbs to their adjectival forms (encoded in the lexicon) if they have one. The reference interpreter will later predicate the verbal relation "give" with "illegal" rather than with the adverbial form "illegally." Since the lexicon links "give" with its nominal form "gift," references of later sentences to "the illegal gift" will select the verbal relation created by the "give" verb in this sentence.

Figure 3-5 presents the deep structure of the same sentence. Only KERN-2 required transformation to deep structure. The deep structure of KERN-2 appears in TRANSFORM-1. Figure 3-5 presents the same parse graph as does Figure 3-4, except that it masks out the subsidiary phrasal structures of KERN-2. Figure 3-4 likewise masks out TRANSFORM-1 and its subsidiary structure. Because it has an indirect object, KERN-2 required the dative transformation. There are two types of dative verbs in English, those whose indirect objects are transformed into objects of the preposition "for" (e.g., "buy," "kill") and those whose indirect objects become objects of the preposition "to" (e.g., "sell," "give"). Whenever a particular verb is first seen this
information is acquired from users and encached on the verb's lexical
entry.

The constraint poster determines the grammatical relations in any
clause by examining its deep structure representation in the parse
graph. Since the dative transformation ensures that no more than one
object will appear for any verb, only one object will be created for
each verbal relation in the knowledge representation. Accusative
intransitive verbs (e.g., "go," "wait") have no object at all. For
these cases, the reference interpreter (by convention) makes the
relation point to its subject as its object as well. As a result, all
verbal relations have just one subject and just one object. Any
operation performed on the knowledge representation can rely on this
fact. These operations are simplified as a consequence.

Figure 3-6 presents the surface structure for the sentence "The
Contras were supported by North because Reagan gave him the impression
that he wanted to destabilize the Nicaraguan regime and because he
served on the NSC staff." This sentence is quite complex, as is its
derivation in Figure 3-6. Note that there are two coreference cycles
in this sentence -- one for "Reagan" and one for "North." The "Reagan"
coreferences are the result of both anaphora resolution and nominal
lowerings while the North cycle is due entirely to anaphora resolution.
The "Reagan" in NOMINAL-6 and KERN-3 was originally the pronoun "he".
Since the clausal attachment code runs before the anaphora resolution

---

28 All relations in the knowledge representation, including
prepositional relations and dummy relations that denote one-place
properties, have just one subject and one object.
code (so that anaphora resolution can exploit the clausal structure of the sentence), the noun phrase that was lowered into the subject position (NOMINAL-8) of KERN-4 was also the pronominal "he". When the "he" of NOMINAL-6 was resolved to be the "Reagan" of NOMINAL-3, the coreference pointer to NOMINAL-8 was thus present, and NOMINAL-8 was also converted to REAGAN. Since, at anaphora resolution time, NOMINAL-6 already pointed to NOMINAL-8 as its coreferent and vice versa, NOMINAL-3 simply changed NOMINAL-6's pointer to itself and made itself point to NOMINAL-8, thereby establishing the coreference cycle.

KERN-2 and KERN-5 are conjoined adverbial clauses embedded in JUNCTIVE-1 under ADVERBIAL-1 in the matrix sentence. When the clausal attachment code encountered the complement agent "and because" (now encached on KERN-5 but not shown in the figure), it searched up the clausal hierarchy from the (linearly) preceding clause KERN-4 in an effort to find the best clause with which to conjoin. Since KERN-2 shared the complement "because", it decided that this was the best clause with which to conjoin and spliced the juntive into the graph below the adverbial that had, until the, directly encached KERN-2 on its adverb slot.

Figure 3-7 presents the deep-structure representation of this sentence. Two clauses were transformed. The matrix sentence is passive and was therefore converted to the active voice in TRANSFORM-1. KERN-2, "Reagan gave the KERN-3 impression to North", underwent the passive transformation, resulting in TRANSFORM-2. Notice that the object-orientation of the parse graphs simplified these transformations. There is no need for a great amount of tree-trimming.
Instead, the transformations simply copy and reshuffle the order of the phrasal agents. The passive transformation also removes the "by" preposition from its deep-structure subject.

Before moving on, notice that, inside KERN-5, NOMINAL-13 is embedded as an adjective within NOMINAL-12, "on the NSC staff", in both Figures 3-6 and 3-7. Thus, NOMINAL-13 has a representation that matches the representation of "on the staff of the NSC", except that the preposition is "+of+" instead of "of". If the prepositional phrase had been "on the NSC's staff", the same representation would have resulted, except that the preposition would have been "*of*". The constraint poster converts these special prepositions to "of" and adds a constraint indicating that the relationship is genitive ("+of+") or possessive ("*of*"). The reference interpreter then creates the same network structures for each in the knowledge representation, except that the "of" relation is predicated as being genitive or possessive, as appropriate. This canonicalization ensures that any references of "North served on the staff of the NSC", "North served on the NSC's staff", or "North served on the NSC staff" will select the same "serve" relation.

A similar canonicalization appears in NOMINAL-8 of KERN-4, in which "the Nicaraguan regime" is represented as "the regime *of* Nicaragua". The lexicon encaches links between proper nouns (e.g., "Nicaragua", "Canada", "Greece") and their adjectival forms (e.g., "Nicaraguan", Canadian", "Greek") to support this sort of canonicalization. Users specify these links when such words are first encountered.
Figure 3-8 presents the surface-structure parse graph for the sentence "North wanted to sell Iran arms to fund the Contras".

Note that both KERN-1 and KERN-2 have the infinitival complement, "to", but that they are nominalized differently. KERN-1 is the direct object of the "want" verb in the matrix, while KERN-2 is the object of the preposition "for" in the matrix. The difference is ordinarily due to a property of the verb in the matrix clause of the infinitival kern. The lexicon stores a property or "subcategory" of verb roots that specifies whether or not they take infinitival complements. Infinitival expressions following a verb that takes infinitival complements are treated as the direct object of that verb.

Often, such verbs will already have a direct object, as in the sentence "North wanted the Contras to overthrow the Nicaraguan regime", presented in Figure 3-9. As the figure indicates, in such cases the direct object (NOMINAL-2 in Figure 3-9) is deleted from the matrix and lowered into the subject position of the kern.

Whenever an infinitival kern appears after a verb that does not take infinitival complements, or, as in Figure 3-8, when the infinitival kern is preceded by another clausal direct object, it is made the object of a "for" preposition, indicating that the infinitival kern is a purpose for the relationship denoted in the matrix clause. Thus, in the sentence of Figure 3-8, funding the Contras was a purpose behind North's desire to sell arms to Iran (other purposes may of course exist even though they are not stated in the sentence). The reference interpreter will thus create (or find) a "for" relation,
whose subject is the "want" relation of the matrix and whose object is the verbal relation of KERN-2, "fund".

Figure 3-10 presents the deep structure for the sentence "A bible signed by Reagan was presented the Ayatollah by McFarlane". The sentence is rather awkward, but nonetheless grammatical. Its awkwardness is due to the fact that the matrix sentence is both passive and dative, the embedded kern is passive, and "which was" has been deleted between "bible" and "signed" in the surface string. The parser reinserts the deleted "which was" prior to the creation of the phrasal agents. After the clausal attachment routines run, transformations are performed on both the matrix and the embedded kern. The matrix, as indicated in the figure, is thus doubly transformed.

Figure 3-11 presents the deep-structure parse graph for the sentence "When the story broke, North was fired by Reagan". Both the matrix sentence and the embedded kern are transformed. The matrix has undergone the passive transformation while the adverbial kern it embeds has undergone the ergative transformation. Ergative transformations occur whenever a clause's verb is ergative. Ergative verbs are those intransitives whose subject is not the agent of the action, but its object or patient. The transformation makes the subject of the clause its object and makes the token "*something*" the subject. The reference interpreter considers this token to be a variable that ranges over any node in the knowledge base.

Figure 3-12 presents both the surface structure and the deep structure representations of the sentence "It was Hakim who acted as North's banker". The relative clause in this sentence is handled as
before. The matrix "Hakim" noun phrase (NOMINAL-2) attaches the relative, KERN-1, on its adjective slot. A copy of NOMINAL-2 (minus the kern) is then lowered into the kern, in which it assumes the subject role. The relativized object following a copular verb after "it" as a subject indicates the need for the cleft transformation. In this case, the parser moves the relative clause across the copular, so that it now modifies the deep-structure subject, "it". The embedded kern is not copied in deep structure, just moved. Since the transform maintains a map between its phrasal agents and those of its surface structure counterpart (the matrix sentence in this case), NOMINAL-2 in deep structure is known to corefer with NOMINAL-3 in surface structure. Thus, the constraint poster can detect that the copula relates (as an identity) the two "Hakims" in deep structure. These being coreferent, there is no need in the network representation to represent the fact that "Hakim" is identical to itself. Only the embedded relative clause need be created (or found).

Figure 3-13 presents the surface structure and deep structure of the sentence "The special prosecutor discovered that there was much money that was missing". The additional feature that this sentence illustrates is the existential transformation of KERN-1. The transformation move the object across the copula and deletes the existential "there". The grammatical relations that the reference interpreter later sees, therefore, are the deep-structure subject and its constituents, plus the copular verb. Since the copular verb is accusative, the reference interpreter will create (or find) a "be" relation in the knowledge representation that identified the subject as
itself. Since the particular subject will then exist in the knowledge representation, the semantics of the existential expression are captured.
Figure 3-4: A Nominal Lowered Across Two Clausal Boundaries.
Figure 3-5: Deep Structure of the Sentence in Figure 3-4.
Figure 3-7: Deep Structure of the Sentence in Figure 3-6. (continued)
Figure 3-8: Two Ways of Handling Infinitivals.
Figure 3.9: Direct Object Lowered into the Infinitival Clause.
Figure 3-10: A Doubly Transformed Clause.
Figure 3-11: An Ergative inside a Passive.
Figure 3-12: A Cleft Transformation.
Figure 3-13: An Existential Transformation.
3.5 Assessment of the Parsing Model

The advantages of the object-orientation of the parser were outlined in Section 3.1, so they need not be stressed here, except to note again that syntactic transformations are radically simplified when phrasal agents are implemented as conceptual objects. More generally, by implementing the parser as a message-passing environment, the code is radically simplified, facilitating the process of extending the range of the parser's coverage.

The reader would not be left with the impression that all difficulties have magically been eliminated in this parser. In this section, I want to touch on the thornier of these issues as a way of introducing the discussion of linguistic theory in Chapter 4.

3.5.1 Prepositional Phrase Attachment

In discussing the parse graphs above, it was noted that "of" prepositional phrases are attached to the adjective slots of the nominals they modify. Any prepositional phrase that modified a noun phrase (or another prepositional phrase) might be handled in the same way. If a prepositional phrase modifies a verb, it could be handled by attaching it on the adverb slot of an adverbial that modifies the verbal. However, when relying mostly on syntactic information, considerable uncertainty creeps into decisions pertaining to prepositional phrase attachment. For example, consider the following sentences:

(1) Gertrude threw a pie at the man in the car at noon.

(2) Gertrude threw a pie at the man in a fit of rage at his poor humor.
In both sentences, the prepositional phrase "at the man" attaches adverbially to the verb "threw". However, in (1) the prepositional phrase beginning with "in" can attach either adjectively to "man" or adverbially to "threw" (if Gertrude too was in the car), while in (2) it can attach only adverbially to "threw". Further, in (1) "at noon" attaches to adverbially to "threw", but in (2) "at his poor humor" attaches adjectively to "rage".

No syntactic information is present with which to resolve these attachment ambiguities. Nevertheless, we breathing natural language processors are able to analyze these sentences without ambiguity unless they are truly ambiguous, like (1), and our analyses tend to agree intersubjectively. How is this possible? And how can a mechanical natural language processor be made to emulate this capacity?

Obviously, we rely upon semantic and/or pragmatic information to disambiguate such syntactically ambiguous constructions, so it should come as not surprise that proposals exist for simulating this process. Hirst [1984], for example, relies upon what he calls "semantic enquiry" to resolve these and other attachment ambiguities. When faced with such an ambiguity, Hirst recommends a semantic enquiry into the knowledge representation to see how the particular nouns, verbs, and prepositions have previously been attached. Of course, as Hirst himself cautions, this technique is highly fallible. The prior knowledge represented might, for instance, include a statement about the man being in a fit of rage, so that the semantic enquiry would result in an incorrect resolution for (2). Further, the propositions
in the knowledge base might include no information on the subject. Semantic enquiry would be no help at all in that case.

A lexical approach, first advocated by Gruber [1965] and later advanced by Jackendoff [1972] and Gruber [1976], appears to be far more helpful. In essence, the approach would exploit information stored in the lexicon specific to words or morphemes to resolve such attachment ambiguities. Gruber characterizes this sort of information in terms of tree structures, but it appears quite similar to word or morpheme subcategories that have long been postulated as necessary components of a lexical base to control syntactic analyses and generations.

Subcategories are features of particular categories, or parts of speech. For example, a subcategory of a verb might be whether it requires a human agent. "Speak", for instance, might be said to require human agency (but what happens when machines speak?). A subcategory of nouns might specify whether they have spatio-temporal extension. "House" does, but "alacrity" does not. A wide variety of subcategories have been proposed for the various categories.\(^{29}\) In the fully articulated Gruber/Jackendoff model, prepositional attachment ambiguities would be resolved by the subcategories (or "prelexical compositions") of the relevant nouns, prepositions, and verbs.

I will not delve into the approach in any detail here, as that is not my purpose. I will instead describe its usefulness and point out its critical limitation. In other words, I will argue that, although

---

the approach may be helpful and even, perhaps, the best available, it's wrong.

The alternative to exploiting the subcategories of lexically stored tokens would be to encode attachment (or "incorporation" as Gruber would say) constraints on an entry-by-entry basis in the lexicon. This clearly won't do. Restricting ourselves only to cases, like those above, in which a syntactic analyzer must decide whether a preposition attaches adjectivally to a noun or adverbially to a verb, the number of constraints needed for complete coverage of anyone's vocabulary would be:

\[ V \times P \times N^2 \]

where:  
\( V \) is the number of verbs in the language,  
\( P \) is the number of prepositions in the language, and  
\( N \) is the number of nouns in the language.

Literally millions upon millions of entries would be needed. If, however, \( V, P, \) and \( N \) can be limited to the number of subcategories of each of these categories, the number of constraints needed would be reduced to at most a few thousand. This feature provides the proposal with its appeal. To see the difficulty with the approach one must first consider the notion of a subcategory or of "prelexical" structures.

Some subcategories, e.g., the transitivity of verbs and the number and gender of nouns, are arguably syntactic in nature. Others, e.g., the concreteness or abstractness of nouns, the ability of verbs to be performed with an instrument, whether it denotes the motion of an agent, etc., are more semantic than syntactic properties of words or morphemes. They pertain to the meanings of terms. Different senses of
a term may warrant different subcategorial ascriptions. For example, the "stand" at which a vendor sells his wares denotes a physical object with spatio-temporal extension, but a political "stand" does not. To jump right to the point, many subcategories relevant to attachment disambiguation depend crucially upon the discourse context and upon the interpretive presuppositions of an interpreter. Many lexically-based subcategories denote conceptual categories, and, as argued in Chapter 2, no single set of conceptual categories may legitimately be presumed as appropriate for all discourse contexts and for all interpreters.

Another way to state the same point is to echo Eco's observation that a lexicon or dictionary is too restrictive. The lexicon, as construed standardly in linguistics, cannot perform the tasks demanded of it. It is "unable to explain social competence in all its living contradictions, whereas such a competence would have to be acknowledged and explained if a semiotic theory of signification and communication were to be drawn up [Eco, 1976: 98]. What is needed is more akin to an encyclopedia than to a lexicon. Stated generally, we humans are so adept at resolving prepositional attachment ambiguities and other sorts of syntactic ambiguity because we, unlike our computer programs, have life-histories upon which to draw.

If attachment ambiguities are to be resolved by reference to the conceptual structures of the particular terms involved in the ambiguity in an encyclopedic memory, then any disambiguation procedure will require access to a knowledge base of encyclopedic proportions. Hirst, it turns out, is substantially correct, for this is precisely what semantic enquiry does.
This returns us to the interpretive thrust of the Relatus project. If the conceptual structures that guide attachment decisions vary somewhat across interpreters and across discourse contexts, then creating models flexible enough to reflect these variations would be of paramount importance. But there is a bootstrapping problem. In order to acquire the ability to create such models much experimentation with large-scale knowledge representations must be performed, and, in the process of creating these models, attachment ambiguities will arise.

The circle may not be as vicious as it seems. The provisional acceptance of the notion of lexically-based subcategorial attachment disambiguation -- as a temporary simplifying assumption -- may provide the leverage necessary to bootstrap the system to a level of competence at which the simulated life-history of an interpreter can serve as the encyclopedic source for resolving syntactically ambiguous language-processing decisions. This provisional stage may well prove to be rather lengthy, but, because it will allow research to proceed on all fronts, it will also prove itself necessary.

This discussion of attachment ambiguity foreshadows two chapters to follow. Chapter 5 discusses the meaning of the term "conceptual structures", which has been invoked rather vaguely above. Chapter 7 presents the design of an implemented lexicon shell. This shell will serve as the computational framework within which our provisional, lexical stand-in for an encyclopedic life-history will be implemented.
3.5.2 Transformational Grammar

Readers cognizant of the theory of generative or transformational grammar will recognize that the parser described above owes much to the theoretical work performed by Noam Chomsky, his colleagues, and students. Yet readers very familiar with the details of generative theory will recognize that it diverges from it in several crucial places.

For example, the parser creates structures that violate the long-standing linguistic notion of subject-predicate structures. That is, the parser does not embed direct objects, indirect objects, and adverbials underneath verb phrases, creating complex predicate structures. Instead, these structures remain at the clausal level. One justification for this heresy is straightforward: it simplifies transformations. When the objective noun phrases are at the clausal level, transformations like the passive and the dative turn out to be extremely simple. Highly complex tree-pruning operations are avoided, and phrasal movement becomes little more than shuffling the order of a few phrases in a list.

Another heresy is that the parser bucks the contemporary trend in generative grammar away from the construction of explicit deep structures and toward strictly surface interpretations. Of course, this innovation in generative theory was partly necessitated by the overwhelming complexity of transformations, so one might simply say that the parser represents an alternative resolution to the transformational complexity problem. However, the modern
transformationalist would undoubtedly counter that moving direct objects and such to the clausal level complexifies c-commanding.

Essentially, c-commanding refers to a constraint on the embedding of phrasal structures that makes intrasentential coreference resolutions quite straightforward. The anaphora resolutions reflected in the parse graphs presented earlier in this chapter were effected by simulating c-commanding in structures that maintained direct objects and such at the clausal level. It is true that a certain amount of coding gymnastics were required to perform this simulation, but this should not imply that embedded predicate subjects are the correct approach. It makes no sense, after all, to resolve anaphora intrasententially.

Consider two simple sentences.

(3) Elmer like Hortense.

(4) Gertrude knew that Elmer liked her.

When (4) is analyzed in isolation from its discourse context, "her" resolves to "Gertrude". Yet, if (3) precedes it in a discourse, the interpretation of "her" as "Hortense" is far more likely.\(^{30}\) Since anaphora can resolve across sentential boundaries, it makes little sense to perform such resolutions intrasententially. They can be held only provisionally and they are subject to rejection when more satisfying cross-sentential resolutions possibilities appear. Thus, resolution should be delayed until an eidetic representation of the

\(^{30}\)"Gertrude", of course, cannot be ruled out. If this discourse had been verbalized and stress had been placed on "her," an interpretation of "her" as "Gertrude" might well be warranted.
sentence has been constructed in the knowledge base. Intersentential and intrasentential candidates could then be assessed simultaneously.

To be fair, it should be stated that much is to be gained from the study of intrasentential anaphora resolution. Aoun [1985], for example, presents detailed analyses in a contemporary generative approach (GB theory) of intrasentential anaphora resolution phenomena across languages. many of his analyses (e.g., reflexives, parasitic gaps) are relevant only intrasententially. Nevertheless, it should be emphasized that the restrictions of analyses to single sentences isolated from their discourse contexts might blind analysts to intersentential anaphora phenomena that could help explain intrasentential puzzles.

Although syntactic, sentential theory-building in the generative mode has produced a great deal of knowledge about our knowledge of language, at some point soon it must face its limit. The abstraction of syntax from the remainder of language, as many have noted before, simply cannot be justified. Generative theory is grounded in demonstrably false simplifications or idealizations of the language acquisition process that, while opening a large area to inquiry, at the same time impose severe limitations on what can be known about language.

While I want to stress that the generative enterprise has advanced in considerable strides, and may continue to do so for some time, other approaches are feasible that do not as sharply delimit the range of inquiry. In particular, those of us whose research interests range beyond the sharply delimited scope of syntax and into socially relevant
semantic and pragmatic analyses of textual corpora may find it necessary to reject the idealizations that ground generative grammar and pursue alternative lines of investigation.
Chapter 4

The Limitations of Generative Theory

4.1 Introduction

The thesis of nativism, represents the "hard core" or central proposition of Chomsky's generativist research program. The nativism thesis holds that humans are endowed innately with certain principles of "universal grammar" (UG). These principles purportedly account for children's rapid acquisition of their native languages. Generativists consider this a rather remarkable feat, given the barriers to acquisition that children presumably face.

1. The "primary linguistic data" are degenerate. The child's environment includes adult speech with false starts, ungrammaticalities, incomplete expressions, and the like.

2. Ungrammatical exemplars, labeled as such, are not available to children.

3. Natural intelligence does not seem to help. Generally, children acquire language irrespective of their levels of intelligence.

Generative grammar seeks to articulate the presumably innate principles of universal grammar that account for language learning under the conditions of impoverished stimuli.

Led by Chomsky, generative grammarians propound a strongly rationalist thesis. They contend that the child's innate equipment includes an acquisition device that is specially geared for the acquisition of language. In this chapter, I review this position -- the thesis of syntactic exceptionalism. Additionally, I will discuss what I believe are the most potent objections to it.
My purpose is to head off likely objections to the Relatus Parser. These objections will likely arise from those who propose to develop computational models of syntactic performance based closely upon the current theoretical formulations of generative grammar [Berwick and Weinberg, 1984; Barton, 1984]. While the Relatus Parser embodies some of the early, fundamental insights of the generativist school, it does not follow its recent theoretical trends.

For the sake of coherence, I first summarize the generative model of the language "faculty" or "organ" and review the recent shift from a rule-based explanation to an explanation based on principles and parameters. I then contrast two empiricist formulations of language acquisition -- one behaviorist and the other interactionist. The interactionist account of language acquisition turns out to be far more plausible than does the behaviorist account, which cannot withstand Chomsky's critique. Interactionists expose the distorting effects of the generativists' unrealistic idealizations of the language learning situation. The chapter concludes with some observations concerning the enterprise of modeling linguistic performance mechanically and a rationale for the choices adopted when implementing the Relatus Parser.

It should be underscored at the outset that this chapter does not devastate the generative theory. This is not its purpose. In its role as a linguistic theory, the psychologically unrealistic idealizations of generative grammar constitute the sort of simplifying device that often proves useful in theory development. Only when presented as a cognitive theory do these idealizations stretch the bounds of credulity. When pressed dogmatically as the only approach capable of
explaining human linguistic performance and competence -- as
generativists, including Chomsky, often do -- the theory assumes the
color and tenor of ideology.

4.2 Generative Grammar: From Rules to Principles

The transition in generative grammar from rule-based to
principle-based explanations was not motivated by a Wittgensteinian
argument against the possibility of private rules [Kripke, 1982]. Such
an objection might be understood as a threat to the validity of the
individual psychology or "monological" basis of generative grammar.
Quite the contrary, whether rule-based or principle-based, the
explanations of generative grammar imply nothing about the reality of
any entities that might be correlated with the terms used within those
explanations [Chomsky, 1986: 221-275].

There need be no mental rules or principles that correspond to the
rules or principles of such explanations. Just as physicists once used
the terms "atom" and "molecule" without ever having seen either,
linguists may use the terms "rule" and "principle" despite the absence
of perceptual evidence for their reality. Real objects may one day be
discovered to which a term like "rule" or "principle" may be attached,
just as physicists have attached observable entities to the terms
"atom" and "molecule". Until that time, the terms "rule" and
"principle" may be considered metaphors that provide "epistemic access"
[Boyd, 1979] to the knowledge domain of grammar.

A desire to improve the theoretical adequacy of linguistic
explanations motivated the transition from rule-based to
principle-based explanations. Rule-based accounts could describe of
the structures and operations of natural languages, but they could not explain how children universally acquire language. Given the poverty of the stimulus and the cumbersome nature of many of the rules that were posited, the need for a more adequate account was perceived. We will soon return to the notion of theoretical adequacy. First, however, it is necessary to discuss the nature of the transition itself.

4.2.1 Rule-Based Explanations

Until recently, linguistic explanations in generative grammar have been couched in terms of rewrite rules. These rules allow one to compose, decompose, or transform one syntactic structure into another or others. One might for instance decompose a sentence into a noun phrase and a verb phrase, as in

(1). \( S \rightarrow NP + VP \)

The verb phrase might be further decomposed, as in (2),

(2) \( VP \rightarrow V + [NP] \)

where V signifies a verb and the NP in brackets signifies an optional direct object. These phrase structure rules are supplemented by a set of transformation rules, which, on sentence analysis, convert surface syntactic structures to a postulated underlying deep structure in which sentences are represented in a canonical, declarative form.

Throughout its development, the major components of generative theory have remained more or less constant. They are D-structure (deep structure), S-structure (surface structure), PF (phonetic form), and LF (logical form). Under rule-based explanations, the logical
relationships between these components were briefly as follows: D-structure representations were created (base-generated from a lexicon and presumably from some system that represents knowledge) by means of phrase structure rules and converted to S-structure representations by means of transformational rules. S-structures were converted to PF by means of phonetic rules, and by means of other rules to LF, where the scopes of quantifiers are realized.

The levels of PF and LF, in this scheme, constitute the interface between the syntactic component and other linguistic components. At the level of PF the lexical contents receive a phonetic interpretation, while at the level of LF they interact with perceptual systems, production systems, conceptual systems, and pragmatic systems [Chomsky, 1986: 67-68].

In recent years, rule-based explanations have been abandoned in favor of explanations couched in terms of parameterized principles that constitute a child's innate linguistic equipment. Learning one's first language, on this view, involves determining the appropriate values for parameters so that applications of the principles generate only sentences grammatical in that language or in some dialect (or idiolect if the language learner mistakes any parameter values). Under both explanations, children must also learn the subcategory features of particular terms (words or morphemes) encashed in the lexicon. Examples of these features, which, although necessary, remain.

4.2.2 Eliminating Transformational Rules

A crucial early step in the transition from rule-based to principle based explanations was Fiengo's [1974] trace theory. In
earlier formulations, D-structures represented the contents of sentences in a canonical form. The agent of action, for example, is always the subject of a D-structure clause. In sentence generation, a transformation of the D-structure — the passive transformation in this case — might produce a different S-structure in which the agent of action would become the object in a "by" prepositional phrase, while the subject of the sentence would be the object of the action (or patient).

(3) Many books are read by few people.
(4) Few people read many books.

It was noticed that quantifier scopes do not always conserve meaning across transformations between D-structure and S-structure. For example, consider sentence (3) and its active form, (4). The two sentences differ vastly in meaning. Yet, in earlier formulations of generative grammar, (4) was purported to represent the syntactic deep structure of (3). While transformations from S-structure to D-structure canonicalize grammatical relationships without affecting sentence meaning, the same cannot be said for logical relationships such as those of quantifier scope.

Jackendoff [1972: 384-386] proposed that S-structures be retained for logical analysis and that D-structures be retained for analyses of grammatical relations. However, surface-interpretive approaches, such as trace theory and GB theory, do not follow this recommendation. They instead dispense with D-structure altogether, eliminating in the process the transformational rules that map S-structure to and from
D-structure. Rather than create deep structures by applying transformational rules that move sentence constituents, surface interpretivists instead mark the constituents that would be moved with "traces" that are indexed to the location where the constituent would be moved if the D-structure were represented explicitly. Thus, while there is no explicit D-structure, it is represented implicitly in the traces. Marcus [1980] implemented a surface-interpretive parser, representing traces on the property lists of symbols representing constituents to be moved.

Another reason for the abandonment of transformational rules, according to Chomsky [1986: 68-71], is their complexity. The transformational rule for relativization creates a relative clause from a simple expression. One might characterize the rewrite rule as indicated in (5),

\[
(5) \quad NP_1 V NP_2 \rightarrow NP_2 \text{wh-} NP_1 V
\]

where the wh- can be either who or what depending upon a subcategory of the head of the NP proposed to the wh-. For example, consider (6). Rule (5) would relativize (6), forming (7),

\[
(6) \quad \text{Elmer saw the man}
\]

\[
(7) \quad \text{The man [who Elmer saw e]}
\]

where e is an empty category indexed by a trace to the NP, "the man", which has been moved from that position.

The difficulty with this analysis, according to Chomsky, is that too many conditions must be placed on the rule, complexifying it to the
point that it becomes difficult to see how children acquire it without misapplying it (which they do not). First, the wh- phrase can apparently be deleted, producing (8).

(8) The man [Elmer saw e]

Other conditions are also necessary. In some cases, such as (9), the wh- cannot be deleted.\textsuperscript{31} This condition must therefore be incorporated into the rule.

(9) * I wonder [Elmer saw e]

Likewise, (10) is grammatical but (11) is not. This too must be incorporated into the rule.

(10) I wonder [what Elmer found of yours]
(11) * I wonder [who Elmer found of yours]

Like (11), sentences (12), (13), and (14) apply rule (5), producing ungrammatical results. They must therefore also be excluded by placing conditions on the rule.

(12) * I wonder [who a picture of e is on the table]
(13) * the man [[to whom]e I wonder [what\textsubscript{1} Elmer gave e\textsubscript{1} e\textsubscript{2}]]
(14) * what\textsubscript{1} did you meet the man [who e\textsubscript{1} saw e\textsubscript{2}]

It soon became clear that a vast number of conditions would be needed for each rule. We would expect children to utter ungrammatical expressions similar in form to (9) and (11) through (14) above.

\textsuperscript{31}Following standard practice, the expression is marked with an asterisk to indicate its ungrammaticality.
However, children do not make such errors. Why would children not first adopt the rule without conditions and, during a period in which the conditions are acquired, utter such ungrammatical expressions? The anomaly is apprehended another way by noting that the richness of mechanisms posited by this model submits far too many possible grammars to the child for evaluation.

Over the course of several years, from within generative grammar emerged successive proposals offering general principles that, in one way or another, reduced the need for various conditions on transformational rules. These principles were understood to be part of the child's innate linguistic equipment. For the present purpose, these principles and the genesis of the principle-based approach need not be discussed.\textsuperscript{32} It need only be noted that as these principles emerged, transformational movement came to be seen not as the application of rewrite rules, but as the result of the interaction of parameterized principles of UG. The principles, as components of UG, represent the language-independent components of an innate endowment. The parameters, on the other hand, represent features specific to a language. They account for the cross-linguistic differences in passivization, question formation, relativization, etc. Under this principle-based paradigm, the only remaining rule is move-alpha, which states simply that alpha -- any sentence constituent -- is to be moved. Whether and to what location alpha is moved depends entirely on the interactions of principles and the values of their parameters.

\textsuperscript{32}Chomsky [1986: 71-80] provides such a discussion, with references to the appropriate work.
4.2.3 Eliminating Phrase Structure Rules

The elimination of transformational rules would neither make language more learnable nor make the linguistic theory more parsimonious if, in so doing, it increased the complexity of other components. Thus, over the period during which transformation rules were gradually eliminated, an emphasis was also placed on eliminating phrase structure rules, like (1) and (2) above.

The strategy followed was analogous to the strategy for reducing, then eliminating, rules in the transformational component. General principles were adduced which would reduce the variety of rules needed. These principles were then assigned to UG, thereby simplifying the task of first language acquisition.

A lexicon that exists separately from the syntactic component was first posited as a way of reducing the variety of phrase structure rules. Within the lexical entry for a particular word (or morpheme) certain features or "subcategories" of the word would be represented. For example, compare the verbs "hit" and "claim". A feature of "hit" is its inability to take a finite clausal complement, while a feature of "claim" is that it can and that this clausal complement represents a proposition that is claimed. To see this more clearly, consider (15) and (16).

(15) $[\text{NP Elmer}] [\text{VP killed} [\text{AdvP} [\text{S that he might survive}]]$

(16) $[\text{NP Elmer}] [\text{VP claimed} [\text{NP} [\text{S that he might survive}]]$
In (15), the embedded $S$ (sentence), "he might survive", states why Elmer killed, while in (16) it states what Elmer claimed. The syntactic interpretations differ due to differences in the lexically-encoded properties of the heads of the VPs in each. Lexical encoding thus eliminates the need to conditionalize phrase structure rules in order to handle this sort of variation.

A component posited for UG called X-bar theory [Jackendoff, 1977] reduces further the necessity of phrase structure rules. Consider (17), which nominalizes (16).

(17) \[ NP \text{ the } [ N, \text{ claim } [ S \text{ that Elmer might survive} ]] \]

A principle — purportedly universal — of X-bar theory states that $N'$ (if singular) must be preceded by a determiner. The universal principles and language-specific parameters of X-bar theory, then, along with the universal principle of lexical projection, reduce the complexity of rules radically. In fact, phrase-structure rules were so radically de-conditionalized that they were effectively eliminated. The rule move-alpha can thus be generalized to affect-alpha, incorporating the operations not only of the transformational component, but now also of the phrase-structure component.

4.3 Theoretical Adequacy

Now that the transition from rule-based to principle-based explanations in generative grammar has been discussed, we can proceed to a discussion of the theoretical adequacy of the principle-based explanation. Theoretical adequacy in generative linguistics is
generally conceived as progress along the continuum displayed in Figure 4-1 [Chomsky, 1965: 54, 61-62; Wexler and Culicover, 1980: 17-22].

---

| Descriptive Adequacy | Explanatory Adequacy | Feasibility |
---|---|---|

Theoretical Progress

**Figure 4-1: Continuum of Theoretical Adequacy**

The need to progress beyond descriptive adequacy motivated the transition from rule-based to principle-based explanations. Many descriptively adequate grammars that are capable of describing the possible human languages might be fashioned. Logically, the next step would be to constrain the class of grammars to include only those that have explanatory adequacy. These grammars not only cover the possible human grammars (i.e., they are descriptively adequate), they also explain how children can acquire one of these languages on the basis of primary linguistic data — the small sample of sentences to which the child is exposed.

For a linguistic theory to be descriptively adequate, then, it must produce a grammar that is descriptively adequate. For a linguistic theory to be explanatorily adequate, however, it must produce both a descriptively adequate grammar and a learning function
that children use to map the primary linguistic data into a grammar.

Beyond explanatory adequacy lies a third level of theoretical adequacy in linguistics, the level of feasibility. A feasible linguistic theory is both descriptively and explanatorily adequate, but the function that maps from the primary linguistic data to a grammar is further constrained by relevant empirical considerations of processing time and data access. In other words, a linguistic theory cannot be considered feasible if the learning function would require children to access an inordinately large volume of data or if the mapping from primary linguistic data to grammar would consume any more than the relatively few years we know to be consumed.

Generative theory has thus far attained only explanatory adequacy. However, generativists themselves appear to believe that their theoretical formulations currently approach the level of feasibility. By transferring the large quantity of variability from rules that must be acquired to the innate principles of UG, generativists claim they have adequately simplified the work children must perform in acquiring a first language. Children need not learn highly complex, conditionalized rules. The function that maps primary linguistic data into a grammar need only determine the appropriate parameter settings for the innately given principles of UG. This is believed to be a rather straightforward matter, accounting for the rapidity of language acquisition from an impoverished stock of sample sentences. While important details of the properties of UG remain to be specified, the principles-and-parameters approach of modern generative theory explains acquisition given the constraints of feasibility, as formulated.
However, the feasibility criterion, as characterized by generativists, is misformulated. If a linguistic theory is to be deemed reasonably feasible, it should conform to all the properties of the language acquisition situation as discovered by empirical inquiry. A comparison of the account of the language acquisition situation assumed by generativists with the account of that same situation presented by developmental psycholinguists, on the basis of empirical observations, reveals remarkable disparities. As discussed below, these disparities indicate that the primary linguistic data are not nearly as impoverished as the generativists make them out to be.

The principle-based explanations of current generative theory are grounded in fictive idealizations of the language acquisition situation. These idealizations may be seen as simplifying devices that afford the theorist an opportunity to devise explanatory formulations under rather pristine conditions. Once these theoretical formulations have been constructed, the next task would be to relax the idealizations one by one in an effort to attain a feasible theory.

This procedure is by no means unique to linguistics. One could draw many examples of its employment throughout the history of science. A contemporary analogue might be developments in the theory of rational choice in economics and political science. Here, analysts posit unrealistic assumptions about the nature of the choice situation in order to simplify their analyses. Their simplifications include the assumption that actors have access to full information concerning the

33I wish to thank Melissa P. Collie for drawing this analogy to my attention.
nature of the choice situation and the assumption that actors are fully rational\(^{34}\) — assumptions known to be patently false. Rational choice theories, then, enjoy explanatory adequacy insofar as they can predict choice behavior under these conditions. The next step, both for rational choice and generative grammar, would be to successively relax their demonstrably unrealistic psychological assumptions in order to attain feasible theories [Simon, 1985].

Premature practical applications of an infeasible yet explanatorily adequate theory should of course be avoided. Application of the results of rational choice analysis in public policy arenas could prove quite dangerous, especially when one fails to attend to the possible distortions introduced by the simplifying assumptions. Likewise, in investigations of language acquisition, one should avoid the premature jettison of alternatives to generative grammar. As lines of investigation, both rational choice and generative grammar certainly enjoy a great deal of legitimacy, but both have advanced only to an interim stage and neither can reasonably be seen as so indisputable that alternative approaches have become unnecessary.

Chomsky and his students apparently wish to preclude lines of investigation — particularly more empirical approaches — that do not adhere to the received wisdom of generative grammar. A few words from Chomsky on alternative approaches should suffice to demonstrate this.

If the answer to this question of adequacy-in-principle [explanatory adequacy] is positive, . . . we can then turn to

\(^{34}\text{Rational choice analysts typically limit their notion of rationality to means-ends or instrumental rationality, ignoring other forms of social rationality identified, for instance, by Weber.}\)
the question of feasibility: can the inductive procedures (in the empiricist case) or the mechanisms of elaboration and realization of innate schemata (in the rationalist case) succeed in producing grammars within the given constraints of time and access, and within the range of observed uniformity of output? In fact, the second question [feasibility] has rarely been raised in any serious way in connection with empiricist views . . ., since study of the first question [explanatory adequacy] has been sufficient to rule out whatever explicit proposals of an essentially empiricist character have emerged in modern discussions of language acquisition [Chomsky, 1965: 54].

Similar sorts of denials of the likely fruitfulness of empiricist approaches appear in Chomsky's more recent writings, but the quotation above best serves the present purposes. Chomsky apparently fails to see the possibility that his simplifying assumptions (idealizations) of the language acquisition situation so distort the facts of language acquisition that his principle-based account, which is grounded in these idealizations, cannot possibly refute any but the most radically empiricist proposal.

4.4 Empiricist Proposals

4.4.1 Behaviorism

An empiricist proposal that can be ruled out — and Chomsky's own refutation devastates it — is the behaviorist one. Reviewing B. F. Skinner's Verbal Behavior, Chomsky [1959] notes that behaviorist explanations of language learning in the theoretical language of operant conditioning had no grounding in observations of children learning languages. Rather, they were grounded in analogies drawn from empirical research into the behavior of lower species as they learned to achieve rewards (and avoid penalties) through operant conditioning.
Behaviorist research does not bear upon language acquisition, Chomsky argued, because it fails to incorporate the essential contributions to language acquisition provided by (the innate ideas of) the learner. If language learning were merely a matter of reinforcement, to use a memorable example in Chomsky's critique, we would not understand "Your money or your life" unless we had already died many times before.

Behaviorism, of course, is a rather radical form of empiricism. The extremes of Skinnerian behaviorism and Chomskyan rationalism range over a huge expanse. From anywhere between these extremes a proposal for a theory of language acquisition might emerge. Empiricism and rationalism, it should be noted, constitute two poles of a continuum. They are not discrete and mutually incompatible doctrines.

5.4.2 Interactionism

Following a trail first blazed by Vygotsky, developmental psycholinguists of the interactionist school reject as too rationalist the explanations of language acquisition advanced both by Chomsky and by Piaget. Interactionists agree in part with Piaget, who argued for a view that saw language learning as a natural consequence of the development of sensorimotor intelligence grounded in innately provided schemata. Unlike Chomsky, Piaget believed these schemata to be general, not restricted to particular cognitive domains. Interactionists depart from Piaget in granting to the physical and social worlds a far greater role in children's acquisition of

---

35 Chomsky, Piaget and others discuss the relative merits of these positions in [Piatelli-Palmarini, 1980].
language. On the interactionist view, language acquisition involves the interaction of three factors.

1. The child's innate abilities,
2. qualities of the child's world, and
3. qualities of the child's social interactions.

The first of these factors represents the child's contribution to language acquisition. Generative theory prominently features this factor while remaining mute on the other two. As a result, generativists ascribe quite a bit more innate knowledge to the child than may be warranted. Notice that interactionists, unlike radical behaviorists, do not deny the rationalist thesis that some knowledge is innate. To the contrary, interactionists view the innate knowledge of the child as a crucial contributor to language acquisition. The question, again, is a matter of degree. The contributions of the two other factors -- the social and physical worlds -- restrict the scope of language-relevant knowledge that can reasonably be said to be innate.

5.5 Interactionist Objections to Chomskyan Idealizations

Interactionists strongly object to three of Chomsky's idealizations:

1. **The Idealization of Instantaneous Development:**
   Generativists present their theory of language acquisition as synchronic. Generativist arguments for the innateness of particular principles -- even those that are quite complex --

---

36 For an overview of the relationships between these and other approaches to language acquisition, see [Rice and Kemper, 1984].
depend not on bootstrapping from what had previously been learned, but on the perceived absence of any other plausible explanation.

2. **The Monological Idealization:**
   This idealization characterizes language learners as engaged in a rather lonely quest for a personal theory of language. Syntactic hypotheses are considered and tested internally, in isolation from the learner's community.

3. **The Idealization of the Autonomy of Syntax:** Chomsky slights the relevance for language learning of the semantic and pragmatic levels. He wants to carve out syntax as a relatively autonomous linguistic level. Learning a grammar, on his view, is independent of development at the semantic and pragmatic levels.

Let us discuss each in turn.

4.5.1 The Instantaneous Development Idealization

The Instantaneous Development Idealization is clearly false, a point which Chomsky [1975: 119] fully realizes. His belief that this idealization does not too severely distort the facts of language acquisition rests upon the contention that variations in the order in which constructs are presented to the child do not perceptibly impact the child's later communicative abilities [Chomsky, 1975: 121-122]. Schlesinger disputes this.

This argument . . . disregards the ample evidence that adults adapt the way they talk to a child to his level of comprehension . . . , and that, should they fail to do so, the child seems able to filter out speech that is too difficult for him . . . . Certainly, the child may encounter [sentences] . . . in any order as speech directed at adults, and if he would try to process this he might get some queer ideas about language -- but who listens? [Schlesinger, 1982: 33].

The child's ability to perceive selectively completely subverts the instantaneous development idealization. As a result, the
innateness of the more complex principles posited for UG are called into question. On the interactionist view, children bootstrap their knowledge of language in a predictable (if not definite) order. If this is indeed the case, then generative contentions of the innateness of principles governing the derivations of highly complex sentences seem far less feasible. Certainly they cannot be considered to have been demonstrated to be innate.

In his computer simulation of language acquisition, Berwick [1985: 30] pursues the consequences of a developmental approach to language acquisition. Relaxing the instantaneous development idealization, he asserts that there is no necessary contradiction between the instantaneous and non-instantaneous viewpoints, but only a difference in research objectives. The instantaneous fiction is meant only to simplify the task of describing the grammar, and it must be relaxed in order to examine the actual course of language acquisition. On the basis of his simulations and from arguments grounded in Keil's [1979] theory of conceptual development (see Section 4.6.2 below) and evidence from the study of aphasia (see Section 4.6.3 below), Berwick [1985: 309] argues for the superiority of non-instantaneous development models.

In passing it should be noted that, although Berwick relaxes the instantaneous development idealization, other idealizations (notably the monological idealization and an assumption that minimizes semantic sources of syntactic learning) remain in force. Additionally, a lexicon and lexical features, thematic role assignment principles, and parts of x-bar theory are fixed in his model [Berwick, 1985: 49-51]. Thus,
while the relaxation of the instantaneous development idealization indeed represents an advance and makes the model interesting, Berwick offers no evidence for or against the generativist versus the interactionist model of language acquisition.

In any event, the preference for non-instantaneous models should cast some doubt upon the purported universality of the more complex principles postulated for UG. An additional principle of UG could stipulate the order in which principles are effectuated or in which parameters become settable, but this would stretch an already incredulous theory much further toward the outer limits of credulity. More plausibly, the child's developmental interactions within a physical and social environment open the possibility at a particular level of development for the acquisition qua induction of linguistic rules, which in turn will provide the child an opportunity to induce additional rules at later stages. Any so-called "ordering principle" would be epiphenomenal on this account.

4.5.2 The Monological Idealization

Chomsky characterizes the language learner as a solitary seeker of a theory of language who is confronted with the seemingly insuperable task of developing a personal theory of a language from a degenerate stock of primary linguistic data. The data are degenerate in the sense that they are replete with false starts, incomplete and ill-formed sentences, and the like.

Empirical observations of the language acquisition situation, however, tend to confirm an entirely opposite characterization. Again,
Schlesinger [1982: 37] cites a host of studies that falsify this assumption. The evidence he reviews indicates the following:

1. Adults speak more slowly to children and pause more frequently to allow children to analyze the utterances.

2. Adults gear the complexity of their speech to the child's level of comprehension. Sentences are short, simple, and rarely grammatically incorrect.

3. Children as young as four have been observed adjusting their speech patterns when speaking to younger children.

4. The responses of children to an adult's efforts to communicate appear to cue adults whether and when to slow down the rates of their speech or to simplify their utterances.

Children, in short, are not the passive observers of adult linguistic interactions that the generativists make them out to be. Children want to learn language in order to communicate. Adults and older children facilitate this process by providing a core stock of non-degenerate primary linguistic data from which the child can generalize.

Generativists must deny these results in order for the argument for innateness from the poverty of the stimulus to hold water. Most often the results are not denied. They are simply ignored. Among generativists, Wexler and Culicover [1980: 78] are more circumspect. Concluding their critical analysis of the literature on "Motherese" (language as used by mothers to their pre-verbal children) and "Baby Talk", they write:

[We do not claim that conditions of input have nothing to do with language acquisition, only that such conditions cannot play a role in the acquisition of language so central that special structural principles are not needed.
Again, the difference between rationalist and empiricist approaches to language acquisition is continuous, not discrete. Few modern observers, if any, would posit a *tabula rasa* theory of language acquisition. The question at hand does not concern whether special (i.e. innate) structural principles are needed. Surely they are. The question concerns the scope of innate principles. Wexler and Culicover continue:

In our view, the only way to discover the exact role that such input conditions play is to create a theory of language learning that can account for the fact that language is learned, and to incorporate within this theory empirically acceptable assumptions about the role of input. Short of such a theory, precise arguments simply cannot be made about how particular environmental conditions function in language acquisition.

Here, Wexler and Culicover are right on target. But the generativist theory of language acquisition they proceed to discuss for the next 400 pages or so does not incorporate empirically acceptable assumptions about the role of the input.

4.5.3 The Autonomy Idealization

Chomsky [1975: 31-32] cites as evidence for the innateness of linguistic knowledge the way that questions are produced from declaratives. For example, consider (18-22).

(18) The man is in the room.

(19) Is the man in the room?

(20) The man who is tall is in the room.

(21) Is the man who is tall in the room?
(22) * Is the man who tall is in the room?

The simplest generalization of question formation from simple sentences like (18) might be "move the verb to the front". Such a rule would, for example, correctly produce (19) as the interrogative form of (18). But notice that this rule would incorrectly produce the ungrammatical (22) from (20) instead of the grammatical (21). This, according to Chomsky, counts as evidence that UG contains the principle that all syntactic rules are structure-dependent. The child forms the hypothesis not that the first verb is fronted, but that the verb paired with the subject of the sentence (the second "is" in (20)) is fronted.

An alternative explanation might suppose that the structures of syntax are acquired rather than innate. A more plausible explanation than either, however, holds that this sort of structure is indeed innate, but is not specific to syntax nor even to language in general.

Prior to a child's first production of an intelligible linguistic utterance (apart from cries, giggling, etc.), he or she has already learned quite a bit. Balls roll when pushed, rattles make noises when shaken, etc. The pre-verbal child, therefore, already knows that agents can act upon objects. These universal pre-verbal categories — agent, action, object (or patient) — are semantic, not syntactic.\(^{37}\) Nevertheless, they provide the ground from which children derive the

\(^{37}\) Occasionally, the claim is made that Whorf [1956] discovered that these categories vary across languages and thus cannot be universal. Whorf, however, shows only that these categories receive non-instrumental interpretations in the American Indian languages he surveyed.
grammatical categories that constitute the building blocks for syntactic structures.

During the course of the preparation for meals, during other caregiving activities, and with every locomotion, the mother performs many acts that provide models for the subject action structure. When these activities have a direct and visible impact upon an object, the infant's observation must lead to the more complex mental structure of subject-action-object. Examples of such structures would be: the mother (subject) carrying (action) a bottle (object); the mother (subject) bringing (action) a pillow (object) [Moerk, 1977: 97].

Evidence for the pre-verbal child's acquisition of the semantic agent-action-object triad prior to any evidence that syntactic knowledge has been acquired emerges from studies of children's early single-word usage. In general, researchers agree that these children use single words "holophrastically". That is, with a single word they refer to a whole relational configuration. For example, a child's utterance of the word "ball" can mean "the ball is there", "throw me the ball", "where is the ball?", or any one of many such intentions. That a child refers to something other than simply the empirical referent of the word is evident from his or her gestures and facial expressions [Rodgon, 1976; Ingram, 1979].

The agent-action-object structure is also present in the speech act. When you speak to me, I recognize you as an agent taking an action (speaking) directed at me. Pylyshyn [1977] notes that the development of linguistic competence presupposes (as a transcendental necessity) the recognition of speech acts. The child would have no reason whatsoever to even attempt the acquisition of a language unless speech acts are already recognized as such.
Beyond the agent-action-object triad, other structural features already exist in the structure of reality presented to the pre-verbal child. Objects are already locatable by children before the acquisition of expressions (e.g., prepositional phrases) used to denote phrases. A child knows that a ball is on the floor long before he or she has acquired the ability to use the prepositional phrase "on the floor" to express the location of the ball.

Caramazza and Zurif [1978] conducted an experiment that directly tests the proposition that syntactic knowledge is bootstrapped from semantic and pragmatic knowledge. They examined the performance of children of various ages on a task that assessed the relative usage of syntactic, semantic, and pragmatic operations. They compared the relative abilities of children to identify pictures corresponding to center-embedded sentence constructions (i.e., "The cat that the dog is chasing is brown"), some of which included semantic and pragmatic constraints that would aid identification and others of which did not. They found significant differences between the performance of children from ages 3 to 5 and the performance of older children.

... [W]hatever strategies the 3 to 5 year olds may be using in decoding the meaning of sentences, it is certainly not an algorithmic-like syntactic processing device. Indeed, it seems that up to age 5 subjects are relying heavily on diagnostic indicators of a semantic nature to help them make their choice. It is only at age 6 ... , that the child begins to use strategies that reflect the importance of syntactic cues (pp. 152-153).

For Schlesinger [1982] and most other developmental psycholinguists, the universal structure of experience -- in this case the universal fact that objects have locations -- accounts for the ease
with which children acquire syntactic structures. Reality, on this view, provides a template that children use into which to map the expressions of their native tongues [Sinclair, 1971; Clark, 1973; Lempert and Kinsbourne, 1979].

This account can receive either of two interpretations. An empiricist can believe that these environmental structures are acquired experientially by the child, while a rationalist can believe that children impose these structures on the environment. Schlesinger himself apparently believes the empiricist version, while others, most notably Bierwisch [1967], would attribute the correlation between the perceptual field and linguistic structures to innate predispositions. On neither interpretation, however, can one account for the structural relationships that inhere in language by reference to an exclusively syntactic universal grammar.

Interactionist views with respect to the contribution of the environment are entirely consonant with the Piagetian approach to language acquisition. Piaget, following up on Merleau-Ponty's [1962] gestural approach to language acquisition, held that the pre-verbal development of sensorimotor schemata fostered the child's acquisition of language. Interactionists understand these sensorimotor schemata as the aboriginal templates for semantic schemata of agent, action, and patient, which in turn serve as templates for the grammatical categories.

Chomsky bases his skepticism of Piaget's approach entirely upon the fact that no full-blown theory of language acquisition has emerged
from it. Interactionists, on the other hand, do not reject Piaget's account so much as they extend it, incorporating the social interactions of the child as essential elements in any feasible theory of language acquisition [Rice and Kemper, 1984: 31]. While Chomsky's skepticism of Piaget's underspecified approach may certainly be warranted, more recent and more well-specified interactionist accounts are beginning to emerge (e.g., [Schlesinger, 1982]). And as they emerge, generativists will become increasingly hard-pressed to defend their radical rationalism.

4.6 Against Syntactic Exceptionalism

4.6.1 The Omission of Knowledge Representation

Chomsky's model of the relations between the major components of language, outlined in Section 4.2.1 above, omits discussion of a critical component, although he alludes to it in passing when discussing the level of logical form (LF). LF, according to Chomsky [1986: 67-68], interacts with perceptual systems, production systems, conceptual systems, and pragmatic systems. These systems remain unspecified in Chomsky's model, but they all fall under the general rubric of "knowledge representation," taking "production systems" and "pragmatic systems" to refer to systems of operations that utilize knowledge representations as data.

As the interactionists remind us, children have already acquired a certain amount of knowledge before the begin to evidence knowledge of a

---

38 The Chomsky-Piaget debates, again, are presented in [Piatelli-Palmarini, 1980].
language. This knowledge is presumably encoded in this knowledge representation to which language finds an interface through the level of logical form. Because Chomsky presents no theory of knowledge representation to accompany his syntactic theory, he blinds himself to the possibility that many of the principles of UG, which he considers to be purely syntactic, originate in the knowledge representation system.

On this view, languages have universal characteristics not because we are endowed universally with a special syntactic faculty, but because all our experiences must be represented within an innate system for representing knowledge. Learning, on this view, takes place within this knowledge representation system.

Frame-based knowledge representation systems constitute computational theories of the device Chomsky omits from his model of language. The triadic structures of such systems and their compositionality, as described in Chapter 1, support the acquisition of the structural relationships of syntax in a rather straightforward manner. The subject-relation-object sequence of both action schemata and utterances are easily mapped into the triadic structures of these systems, and their compositionality guarantees the representability of additional relations that link subjects, relations, and objects to other lexical contents. The acquisition of syntax, then, can be explained without reference to any specifically syntactic innate mechanisms.

---

39 One place properties are represented in such systems by linking the property to its object by means of a dummy relation.
Another negative externality of the omission of knowledge representation from Chomsky's account concerns the problem of reference. The time-complexity of reference, or the mapping of the contents of a sentence to a knowledge representation, must be sublinear in the size of the knowledge representation. Simply put, it should not take longer to understand a sentence when the understander knows more. Any theory that implies that larger amounts of knowledge impede the understanding process should certainly be viewed with suspicion.

Mallery's [1987] algorithm for referencing into Gnoscoere by interpreting constraints shows that this class of knowledge representation system supports sub-linear reference. Duffy and Mallery [1984] argue that explicit deep structures are crucial to the process of referencing the contents of syntactic structures consistently and efficiently into such a knowledge representation. They challenged the generativists who advocate surface-interpretive grammatical theories to show how consistent and efficient reference can be performed without explicit deep structures.

Perhaps the assumption of many generativists of the modular character of components of the language faculty leads them to assume that any interface between the syntactic component and a knowledge representation system will be straightforward. But this is something to be demonstrated, not merely assumed. In any event, the absence of any knowledge representation theory whatever constitutes a critical limitation for the applicability of contemporary generative theory to cognitive modeling.
The foregoing shows how one of the unrealistic idealizations of generative grammar prevents its feasibility as a theory of language. This is the idealization of the autonomy of syntax. The boundaries between the linguistic levels of syntactics, semantics, and pragmatics are exceedingly fuzzy on anyone's interpretation. Each tends to shade into one another.

4.6.2 Hidden Semantics: The Lexicon

Even within contemporary generative grammar, semantic knowledge tends to intrude. Much of the intrusion is hidden in "the lexicon"--a purported component of language. Most modern linguistic theories, generative grammar among them, are lexically based. This means that they posit a central role to the lexicon. In Chomskyan terminology, terms are projected from the lexicon with subcategorization frames. For example consider sentences (23) and (24).

(23) The astrologer would say the moon is green.

(24) * The doorknob would say the moon is green.

(23) is well-formed because "astrologer" has the subcategorial lexical property [+ human], which conforms to a subcategorial restriction on the verb "say" that requires its subject be [+ human]. (24), meanwhile, is ill-formed because "doorknob" has the subcategorial lexical property [- human], and is thus incompatible with "say". Under ordinary circumstances (that is, apart from fantasy speech contexts), then, sentences like (24) would not be produced.

The notion of lexically based sentence production and understanding is quite appealing. For AI natural language programs,
such a data structure can efficiently store subcategorial selection restrictions. In fact, a later chapter of this thesis describes the GEL Shell — a generic and extensible lexicon shell — which is designed to serve just this purpose.

It is important to remember, however, that lexical databases serve only as a computational convenience. There exists no evidence whatever that people maintain anything akin to a mental lexicon. Far more plausibly, when we wish to know whether a particular noun can serve as the subject of a particular verb or when we wish to make similar selections for syntactic purposes, we consult not a special-purpose dictionary, but our experiential knowledge of that noun and its exemplars as represented in an encyclopedic general-purpose memory [Eco, 1976: 105-121]. Like the humanness or non-humanness of the astrologer and the doorknob above, many (but not all) subcategorial selection restrictions are semantic or pragmatic, but not syntactic, in nature.

Keil [1979] develops and tests a cognitive theory of category development based upon Sommers' [1971] theory of predicability and ontological categories. Sommers' approach is best illustrated by reference to two isomorphic tree structures. The trees, adapted by Keil [1979: 15-16], are displayed in Figure 4.2
Figure 4-2: Sommers' Ontological and Predicability Trees
Many of the selection restrictions posited for the lexicon follow naturally from the predicability of members of the ontological categories. To the extent that Keil's application of Sommers' theory to conceptual development is successful, the generativist belief in the relative autonomy of syntactic knowledge is shaken further. Learning these (or similar) ontological categories becomes a precondition for the development of successful verbal practice.

The proposed theory explains several seemingly diverse phenomena with one unified account. Other work has generally treated phenomena such as selection restrictions as isolated cases and has failed to see the links to other phenomena, such as the link between similarity and natural classes via a common underlying knowledge base [Keil, 1979: 22].

If learning natural classes, or the development of "a common underlying knowledge base" catalyzes the child's ability to utter comprehensible sentences, is it not possible that the child progressively deuto-learns or learns how to learn better [Argyris and Schon, 1978]? If, in their interactions with their physical and social worlds, children can learn to learn better, then is it not also

---

40 Keil [1979: 16] believes that some semantic selection restrictions do not rely upon ontological categories. He would restrict the restrictions to those that Ryle and others term "category mistakes". For example, he states that ontological categories explain the ill-formedness of "the idea is green", but not the ill-formedness of "the bachelor is married". It seems possible, however, that as one's ontological knowledge becomes more elaborated such mistakes can similarly be covered. That is, mistakes like "the bachelor is married" might be covered by ontological categories as the right hand side of the node marked "IS HONEST, IS SORRY" in the predicability tree and marked "SENTIENT BEINGS" in the ontological tree becomes elaborated. The predicate "IS MARRIED" might then apply to "man" and "girl" but not to other categories of sentient being, like "bachelor" and "spinster".
possible that learning rules of syntax can be a rather straightforward matter for the child?

4.6.3 Neuroscientific Evidence

The evidence from neuroscience in no way confirms the model of syntactic exceptionalism. The observation that particular regions of the brain are devoted to syntactic processing is consistent with syntactic exceptionalism. However, it is also consistent with the interactionist view that most syntactic abilities emerge from the child's semantic and pragmatic abilities that develop in interaction with his or her physical and social worlds. Syntactic brain localization may itself develop.

There is evidence that the degeneration of the language abilities of adult aphasics\textsuperscript{41} represents a regression to the pre-syntactic processing level observed in young children. Caramazza and Zurif [1978] assessed the performance of adult aphasics on the task described in section 4.5.3 above and found that they regressed to a level of competence that relied on semantic and pragmatic knowledge and evidenced no use of algorithm-like syntactic processing. The performance of aphasics did not correspond precisely to the performance of the 3 to 5 year old children, however, contrary to the regression hypothesis originally postulated by Jakobsen [1968].

\textsuperscript{41}Aphasia, a condition caused by a lesion in the left hemisphere of the temporal cortex, diminishes language capacity. Childhood aphasia can be completely overcome. Child aphasics transfer the lost abilities to the right hemisphere. Teens and adults who become aphasic can improve, but only marginally. Presumably, according to Penfield [1971: 329], teens and adults have established a "neuronal set" that cannot easily be overcome and have allocated regions in the non-dominant hemisphere to other tasks.
They conclude that Jakobsen was essentially correct, but that the notion of aphasic regression requires slight modification. They believe the Piagetian notion of a hierarchy of historical gestalts, in which at successive levels of development language learners modify earlier heuristics while learning new ones [Piaget, 1963]. Thus, while aphasic regression does occur, later modifications to early language abilities are retained. For Caramazza and Zurif,

... language comprehension is to be seen as the result of a complex inferential process. The rules can be separated roughly into semantic-like and syntactic-like processes. Both kinds of inference rules operate at all levels of functioning, whether it be at the level of the young child, the adult, or the aphasic. The difference is to be found predominantly in the relative distribution of the use of each of these two sets of rules. The young child first develops semantic inferential rules consonant with his general cognitive development; as his general cognitive apparatus expands so do his semantic inferential strategies. In addition to the development of these semantic strategies, the child constructs a more rigorous inference system based on the syntactic regularities of his language. In the case of language breakdown, if syntactic knowledge is impaired the aphasic will have to rely primarily on his semantic inference system, a system that includes a lifetime of experience and knowledge. It seems unlikely that the inference processes used by the aphasic will strictly correspond to those used by the child. There will certainly be a similarity in performance, but, consonant with the hierarchical model, there is no point where complete correspondence obtains between the child's developing rules for mapping meaning and the residual functions the aphasic is using to infer meanings (p. 160).

These results, while consonant with an interactionist model of language acquisition that recognizes that syntactic, semantic, and pragmatic abilities bootstrap each other, are incomprehensible within the model of syntactic exceptionalism. If children learn language by fixing the parameters of innate principles of universal grammar without reference to the earlier development of syntactic and pragmatic
abilities, why do younger children rely exclusively on semantic and pragmatic cues in the center-embedded sentence tasks while only older children are able to utilize syntactic knowledge to match sentences to pictures? Further, why do aphasics regress to a stage where they utilize semantic and pragmatic cues? Why don't they instead regress to a stage where only purportedly innate syntactic cues are available? The answer, of course, is that there is no such stage, for there are no innate, specifically syntactic, mechanisms.

4.6.4 Implications of Radical Rationalism

No claim is made for the comprehensiveness of any of the preceding treatments. Each topic in this section itself deserves comprehensive treatment, but that is left for another day or perhaps for other researchers more conversant with each topic. My aim has only been to illustrate various reasons why I believe the thesis of syntactic exceptionalism to be incoherent. Certainly, the thesis is internally coherent, but it simply does not cohere with the facts of language acquisition.

The generative approach to language, it should be remembered, is not at issue here. Insofar as linguists are concerned to describe adequately the facts of syntax across human languages, generative theory has much to say, and it represents a remarkable advance in our understanding of language. However, when elevated beyond a description of the abstract competence of an ideal user of a language and pitched as a theory of the acquisition of language, generative theories cannot be accepted unless they square with the empirical facts of language acquisition. The principles-and-parameters approach does not.
4.6.4.1 Research Implications

Schlesinger [1982: 28] recommends an approach to research in language acquisition that would incorporate the insights that linguists are able to provide, but would reject the unfeasible approach to acquisition that Chomsky offers. This research program would consist of three components.

1. Linguistic research that aims to discover generalizations about the principles that underly human language.
2. Psychological research that aims to discover how the principles revealed in (1) operate in other cognitive domains.
3. Psychological research that constructs and tests explanations of how the principles in (1) develop through interaction with the environment.

Unless adopts the Chomsky's syntactic exceptionalism, this research strategy seems quite reasonable. However, Chomsky consistently denies the possibility that (2) and (3) can provide any insights. He instead wishes simply to deduce the principles of a universal grammar that, he believes, can alone account for the "facts" of language acquisition as he sees them. Schlesinger [1982: 39] comments on Chomsky's intransigence.

What Chomsky wants to do is establish Rationalism by default, for lack of an alternative theory. He argues against the prospects of an approach that would trace the acquisition of linguistic principles through experience with language and solely on the basis of general cognitive strategies. However, ... his arguments rest on faulty factual foundations, and are partly launched against a particular, clearly inefficient, type of empiricist explanation [behaviorism]. This does not imply, of course, that the alternative approach proposed [in the research program proposed above] ... has been vindicated.
Obviously, psychological research in those areas has a long way to go; in fact, it has barely started, and it is still far from engendering a well-substantiated theory. Chomsky's writings on this topic aim to bury such a theory prenatally.

As a theory of language acquisition, generative theory is degenerating. The hard core of generative theory, the innateness of specifically syntactic universal principles, is under sharp attack. Chomsky's reactive defense of this hard core, while understandable, should not be allowed to stifle the alternative paths of inquiry that emerge from its rejection. The approach Schlesinger urges constitutes a budding research program, which must be sheltered for a while from the condemnations of its powerful, established rival [Lakatos, 1970: 157], so that after reflection on its successes and failures, we can better judge the comparative validity claims of each. To do otherwise would be to engage in self-defeating irrationalism.

4.6.2 Political Implications

Sampson's [1979] libertarian critique of the relation between Chomsky's view of linguistic creativity and his political anarcho-syndicalism misses the point. Chomsky, like Sampson, despises the practice of political domination. Each seeks human emancipation from relations of irrational domination, although they differ vastly in the institutional arrangements they propose for ensuring liberty. Sampson would free the inventive genius of humankind from the restrictions of paternalistic governmental regulation, while Chomsky would free the productive skills of humankind from the domination of the employer, organizing work along the lines of voluntary associations of worker-managers.
Habermas [1979] reveals a normative moment in the idea of communicative competence. Habermas postulates that in the very act of engaging in discourse with others, we presume the possibility of coming to a rational consensus. By communicating we declare rational consensus to be in our emancipatory interest. We posit the possibility of approaching the point at which our community becomes an ideal communication community — a community in which all its members have communicative competence — the ability to enter into discourse unfettered both by the external constraints of coercive domination and by internalized psychological constraints resulting from relations of domination. In the ideal communication community, then, decisions are based upon the force of the stronger argument rather than the force of power.

Chomsky's radical rationalism partly de-problematizes and de-thematizes the idea of attaining the communication community. If the attainment of linguistic competence is simply a matter of assigning values to the parameters of principles genetically provided, then we need not overly concern ourselves with ensuring that members of the community attain the training necessary to become effective communicators. Certainly, with or without a universal grammar the effects of domination may inhibit free and effective discourse. But on the presumptions of Chomsky's radical rationalism, we need not concern ourselves with the possibility that inequitable social conditions may foster inequities in the communicative competence of entire sectors of the community. Communicative competence, on the Chomskyan view of language, would be guaranteed as an essential component of human
nature. Only later — after the primary language is acquired — can it be distorted.

Sampson perceives a basic consistency between Chomsky's linguistics and Chomsky's politics and argues that both are mistaken. However, from the perspective of critical theory, of which Habermas and Apel are two major contemporary luminaries, Chomsky's linguistics contradicts his politics, and critical theorists would find wanting the linguistics more than the politics.

4.7 The Transformation of Linguistics

4.7.1 A Return to Rules

Once the monological idealization of generative grammar is recognized as a falsehood, the child's acquisition of grammar can no longer be viewed as the simple plugging-in of values for the parameters of innate principles. The child instead is now seen as internalizing the linguistic norms of the community. That is, the child progressively internalizes as norms the rules used in the community for expressing oneself and the relations perceived in one's environs. The child develops a theory of the language, expressible (not by the child, but perhaps ultimately by the scientist) as the totality of those rules.

As mentioned in section 4.2 above, Chomsky denies that the Wittgensteinian argument concerning the social nature of rules threaten the "individual psychology" approach of generative theory, since "rules" are only a convenient nomenclature for which no psychological reality has ever been postulated. Recall that Chomsky's defense was
granted in that section. We have since discovered, however, that the facts of the language acquisition situation itself provide the necessary grounds for dismissing the "individual psychology" basis of generative theory. Once language acquisition is seen as the internalization of social norms, the Wittgensteinian conditions under which we may ascribe rule-following behavior to language users are met. Linguistics is (as it was before Chomsky) a social science, not a natural science.

With his theoretical reliance on innateness and his treatment of an explicitly syntactic "language acquisition mechanism" as a human organ on a par with the liver or the kidneys, Chomsky wants to categorize linguistics as a branch of biology. The evidence for a biological basis of grammatical competence consists of nothing more than a misspecification of the language acquisition situation and what Putnam [1967] calls a "what else?" argument. Chomsky simply contends that innate linguistics universals can only account for language acquisition. What else could account for it? He is correct, of course. Only innate universals could account for language acquisition if his account of the language acquisition situation were correct. But it isn't.

As emphasized above, certain innate preconditions for language acquisition are indeed necessary. But they need be no more complex, however, than the innate preconditions necessary for the development of human interactive capacities in general. Why do people form groups for collective action? One could certainly claim that people have innate predispositions to form such groups, but this explains nothing. We
require some social explanation for the formation and activities of social groups, and this is something social scientists endeavor to provide. We also require some explanation for the formation and activities of our language capacities. This is something a social science, linguistics, should endeavor to provide.

4.7.2 Linguistics as Social Science

As Apel [1972] notes, Chomsky conceives of the generative enterprise as an empirical science to be developed according to the norms of theory development associated with Popperian critical rationalism. But critical rationalism, like its antecedent, logical empiricism, presumes a strict separation between the subject and the object of science. No such separation is possible in linguistics, as in other social sciences, because the linguist's evidence for corroborating his hypotheses -- the intuitions of native speakers -- already prohibits the separation.

For Apel, linguistics, like other social sciences must be grounded in a hermeneutical-dialectical philosophy of science, in which the object of science is not merely an entity whose behavior is to be analyzed, described, and explained. The object of science is apprehended, on this view, as a partner in communication -- a potential co-subject of science. Linguistics, on Apel's view, becomes a science in which the communicative competence of language users takes precedence over competences, like grammatical competence, that arise only to serve the goals of communication.

Something very important is at issue here. Once we apprehend language acquisition as the internalization of a community's norms for
communicative expression and endeavor to construct theories of communicative competence and, ultimately, communicative performance, linguistics itself is radically transformed. The transformation extends beyond merely a reconceptualization of the domain as a branch of social science. It implies that linguistics becomes a study directed at the very grounds of its knowledge.

The universal linguistic theory can ... be understood as a "meta-theory" of the specific grammars where, according to Chomsky, the latter [a theory of a language] can be understood as possible individual instances of human linguistic competence that are realizable by the child in the language-learning process. If one adopts the latter perspective then the "evaluation measure", which is part of the universal theory, emerges as the objectifying reconstruction of the child's ability ... to select by means of successive construction and self-correction the normatively adequate grammar from the set of possible grammars. This means that the meta-theory has a normative function in the process of theory-formation which the child must perform unconsciously, just as methodology is accorded such a function in the formation of scientific theories. Such an "evaluation measure" is normally — i.e. in the case of natural science — not a part of empirical-analytic theory, but rather it is the task of the normative logic of research as the meta-theory of creative theory-formation [Apel, 1972: 200].

One might, of course, study the acquisition of knowledge in any particular domain in order to inquire into the the process of theory formation, but the acquisition of language presents itself as an especially relevant case. The linguistic norms of a community — and especially the semantic and pragmatic components of those norms — serve a constitutive function for thought. They are, to use Dallmayr's [1984] expression, architectonic of thought. Language provides us with a ready-made roadmap for the more abstract or conceptual aspects of reality not immediately available to the child.
The subsequent acquisition of other theories, whether or not they may be deemed scientific, must cohere with the architecture that language already provides. Novel facts that fail to cohere with the framework that the language provides (i.e., that conflict with linguistically encoded social norms) might thereby be quite difficult to discern. Conversely, whenever a novel fact is discovered and corroborated by extra-linguistic means and to the extent that that fact fails to cohere with the semantic norms of the language, linguistic structures (e.g., the meanings of terms), might themselves require reformulation.

But how can we construct a theory about how we construct theories without already presuming a model of the theory construction process? Put another way, how can we make the epistemological conditions for the validity of scientific constructs itself an object of scientific investigation? This undeniable circularity, presents itself only when language acquisition is apprehended as more than brain science -- as the internalization of the norms of a language community. But it should not be viewed as an insurmountable obstacle to theory formation in this domain. Instead, it should serve to counsel us against the premature, dogmatic advocacy of theoretical constructs, and it should serve to counsel us never to forget the hermeneutical and dialectical nature of the enterprise. Only by remaining willing to submit theoretical proposals, always held only provisionally, to the critical scrutiny of the widest possible discourse community — encompassing not only the community of scholars interested in the subject but also the
much wider community of competent language users — can we prevent the
circularity from leading us radically astray.

4.8 Strategies for Modeling Syntactic Processing

This chapter reflects the understanding of language that informs
the syntactic contributions to the Relatus project. Advocates of no
syntactic processing model can legitimately lay claim to explanatory
adequacy — much less feasibility — unless they can be show that the
model is embedded within a system that can support the acquisition of
knowledge generally and not just syntactic knowledge.

In succeeding chapters, such an acquisition mechanism is
characterized in very broad outline. Of course acquisition of the
acquisition mechanism is a logical impossibility. While we certainly
can learn how to learn better, some basic acquisition device must be
innate. Other principles postulated by generativists for the innate
language "organ" might also be innate, but they need not be seen as
strictly syntactic in nature.

The Projection Principle, for example, holds that the lexical
contents projected at any one linguistic level must also be present at
all other linguistic levels. For example, a subject specified at
S-structure must also appear in D-structure and LF representations.
This principle is not controversial on any account, and may profitably
extended to the eidetic level in the perception of a sentence or of a
configuration in the visual field encoded in a knowledge representation
[Mallery and Duffy, 1987]. Thus the projection principle can be
construed as a general cognitive principle, not an especially
grammatical one.
Likewise, the Subset Principle might also be innate. It holds that in language acquisition children are conservative learners. That is, they order learning hypotheses in such a way that positive examples can disconfirm them.

For many cases, the ordering will force the narrowest possible language to be hypothesized first, so that no alternative target language can be a subset of the hypothesized language. More precisely, no other target language compatible with the triggering data that led to the new hypothesis language can be a proper subset of that language. Importantly, the Subset Principle is necessary and sufficient for acquisition from positive examples [Berwick, 1985: 23].

Children, of course, are presented not only with positive-only (i.e., grammatical) sentences, they are also presented with a positive-only world. They have only this world from which to learn. Not until children reach a relatively advanced level of sophistication can they apprehend the cube-shaped planet Bizarro of DC Comics, where pleasure and pain are reversed, where people cry when happy and laugh when sad, and where Superman performs evil deeds, receiving approval for them. In the actual world in which children learn, the physical behavior of objects and the social behavior of others are appropriate, positive exemplars. To the extent that children experience negative exemplars we might expect learning disabilities to emerge. The Subset Principle, then, like the Projection Principle, is not syntactically exceptional.

Empirical, psychological evidence for the postulated innateness of certain syntactically special principles, like those that concern the locations within syntactic structures at which pronouns can find their
antecedents [Chomsky, 1981: 163], or those that concern the conditions
under which an empty noun phrase can act as a variable [Chomsky, 1981:
175], simply does not exist. These principles are instead deduced from
a theory that fails to cohere with empirical reality. These principles
may certainly contribute to any adequate description of grammatical
phenomena, in which case they have importance for linguistics, but they
do not contribute to an adequate explanation of language acquisition.
Thus, they have minimal value for a cognitively-oriented
psycholinguistics.

Claims to the innateness of particular, rules, principles,
constructs, etc., require some ground if they are to be deemed
credible. Three sorts of grounds for such claims are possible. An
important grounding, of course, is empirical evidence. Developments in
genetics may one day offer this sort of empirical evidence, but as yet
no genetic evidence for the innateness of specifically syntactic
constructs has appeared. One may ground such claims in Kantian
transcendental deductions. That is, they may be based on the argument
that without such a construct empirical experience would not be
possible. Finally, claims may be grounded in deductions from
theory-based assumptions. These, of course, are the most tenuous, for
they may be demolished by demonstrating the falsity of the assumptions.
Chomsky grounds his arguments in just such assumptions, and the
interactionists have, I believe, adequately demonstrated their falsity.

If the Projection Principle and the Subset Principle can be shown
to be necessary principles for acquiring knowledge -- knowledge not
only of language but of the world -- then they may be components of a
general acquisition device. This device, again, cannot itself be acquired, for its acquisition already presumes a pre-existing acquisition device. In other words, these principles (and certainly others) might be groundable in transcendental deductions. The syntactically exceptional principles postulated for UG, however, cannot. Unless claims for these latter principles can be redeemed by incontrovertible empirical evidence, by transcendental deductions, or by deductions from a consensually validated theory of the language acquisition situation, prudence cautions us to regard them very skeptically.

The development of the Relatus Parser, proceeded with full awareness that a parser, by itself, cannot possibly be "cognitively plausible" in any but the weakest sense. Any syntactic model for which one wishes to claim cognitive plausibility or psychological fidelity must be embedded in a system that supports learning, conceived not as the simple binding of the parameters of innately-given principles, but as the abduction and deductive testing of hypotheses, resulting in the progressive elaboration of conceptual schemata.

Mallery's [1987] Gnosocere constitutes the environment in which, for Relatus, this sort of elaboration will occur. However, as explained in Chapter 3, the development of Gnosocere required the implementation of some parsing model. The Relatus Parser was thus implemented as a cognitively implausible model of syntactic processing that can both create syntactic derivations rapidly and model its own transition states in Gnosocere. Once the inferential machinery of Gnosocere is completed, it should be possible to induce, from a large
body of exemplar sentences, rules for syntactic processing. Since Gnosere already has the capacity to model the contents of the texts the parser processes, it should also be possible to incorporate contextual, discourse-related variables into those rules. This project, then seeks to bootstrap a cognitively plausible model of syntactic performance form a cognitively implausible model. It would reverse-engineer syntactic competence.

In order for this project to succeed — in order to extend our abilities to analyze computationally the propositionally representable contents of political texts by simulating the understanding processes of a variety of interpreters who infer meaning on the bases of the words and structures in the text and of the propositionally representable contents of their varied life-histories — mechanisms must be developed for the formation of hypotheses from such computer representations. This is the subject of Chapter 6. However, in order to form hypotheses successfully, we must first state clearly the nature of the conceptual structures from which hypotheses are to be formed. This is the subject of Chapter 5.

Before proceeding, we should note that the discussion of concept formation and hypothesis formation that follow concern more than the development of models of syntactic processing. To be sure, these chapters seek to articulate the "evaluation measure" used to acquire knowledge, syntactic knowledge or otherwise. But in so doing they also articulate, in part at least, what Apel recognizes (cf. Section 4.7 above) recognizes as the "normative logic of research" — the meta-theory of creative theory formation.
Chapter 5
Concept Formation

5.1 Introduction

Interpreters' ascriptions of class or category membership and the conditions used to identify class membership can significantly affect their interpretations. For example, suppose an interpreter — we'll call him Interpreter A — believes that the concept "communists" refers to people who advocate centralized state planning. Suppose further that Interpreter B believes that socialists are people who advocate centralized state planning while communists are socialists who advocate the dictatorship of the proletariat. Suppose further that both interpreters believe that communists seek world domination.

Here we have two distinctly different, yet overlapping, sets of belief. To see how these differences can affect interpretations, imagine that both interpreters learn that the Prime Minister of Sweden advocates centralized state planning. From this fact, Interpreter A might reasonably (albeit incorrectly) infer that the Prime Minister of Sweden also advocates world domination. Interpreter B, on the other hand, would make no such inference.

From this homely example, it should be clear that an AI system designed for general application to political problems should be capable of simulating the varying inferences that result from the differing interpretations that arise from incommensurable attributions of conceptual categories. Because Relatus deploys different knowledge bases for different interpreters, this much is fairly trivial, and
other knowledge systems can be made to perform similarly simply by partitioning knowledge across knowers.

But much more is required. One would like general concepts like "communists" and their intentional attributes to be induced automatically. Another desirable property would be the capacity to reformulate the structure of conceptual hierarchies whenever information incompatible with such a hierarchy is received in enough volume to preclude the rational ignorance reflected in the addition of ceteris paribus conditions associated with concepts. That remark requires clarification, but his must be allayed until section 5.7. One would also like to support the formation of hypotheses from the knowledge encoded in conceptual hierarchies.

Chapter 6 discusses how conceptual hierarchies of the sort Mallery has implemented in Relatus supports the formation of hypotheses, including relatively novel hypotheses. It also presents a simulation in Relatus of a novel hypothesis. For the remainder of this chapter, however, I want to lay the groundwork for the discussion of hypothesis formation by defending against an important recent objection the contemporary formulation of the classical view of concepts and categories which is embodied in Relatus.

Cognitive models that incorporate the concepts and intentions that define those concepts for individuals can certainly aid our understanding of the cognitive incommensurabilities that characterize political conflict situations. But attainment of a coherent and defensible theoretical position with respect to categories and concepts has import for much more than political modeling. Conceptual
categories ground all hypothesis-based learning, including the acquisition of knowledge of one's language and how to use it. Thus, this discussion has relevance for developing models of syntactic acquisition that do not rely upon the innateness of highly articulated syntactic principles. As the discussion in Chapter 4 reflects, this level of knowledge representation — the level at which concepts and hypotheses are formed — more properly serves as the focus of discussions of innateness. Generalized principles of syntax, on this view, are epiphenomena of innate ideas at this deeper level that the children acquire universally as an outcome of their interactions with the physical and social worlds.

Beyond the language of politics and the politics of linguistics, theoretical clarity with respect to conceptualization and hypothesis formation should provide returns applicable indirectly to the concerns of particular sciences. Meta-scientific reflections on the contents of the categories and on causal relations posited between those concepts (i.e., theories) have historically proven invaluable in recognizing and occasionally avoiding pitfalls encountered in the advancement of science. This, of course, is the special role of philosophical analysis. While there exists no general, intersubjective agreement on the contents of concepts, and while the positivists' hope for an ideal description language may be a pipe-dream, it may be possible to reach consensus on the necessarily universal process of conceptualization.

If this consensus can indeed be reached, then computational simulations of the conceptualizations of individual scientists can be constructed and analyzed, as can the discursive, disputational
practices carried on in scientific communities, in which scientists endeavor to articulate, critique, elaborate, and advance the contents of their concepts and theories. The Kuhnian incommensurabilities that characterize competing paradigms could be submitted to members of each competing group. If this would not foster agreement between the competitors, it would at least help make each one intelligible to the other. The general result of such an effort would be the deployment a set of techniques for conducting philosophical analyses that afford the advantage of replicability. The first step, then, is to come to grips with the process of concept formation.

5.2 The Concepts of Concepts

A great deal of experimental results have been produced in recent years concerning the human categorization of natural objects [Mervis and Rosch, 1981]. In a volume that applies these results to the concerns of AI, Smith and Medin [1981] draw conclusions about the nature of human categories and knowledge representation. In particular, they reject the classical, Aristotelian view of categorization in favor of an approach that mixes (fuzzy) probabilistic and Wittgensteinian, exemplar-based views.

Smith and Medin deny some of the fundamental assumptions of many frame-based knowledge representation schemes, including Mallery's Gnoscere [1986], which is used in Relatus.\textsuperscript{42} These systems all incorporate aspects of the classical view of categorization, in which individuals are deemed members of classes when they satisfy the

\textsuperscript{42}See Section I.6.1 of Appendix I for a brief description of Gnoscere.
necessary and sufficient conditions for class membership. Smith and Medin argue that this approach does not cohere with the empirical evidence and is thus untenable.

Below I review the arguments that Smith and Medin offer against the classical view. I find that (a) some of their suppositions about the tenets of the classical simply view do not apply to contemporary formulations of the classical view, and (b) they misinterpret many of the relevant experimental results.

Smith and Medin consider three alternative approaches to conceptual categorization.

1. The classical view characterizes concepts as summary descriptions of the features deemed singly necessary and jointly sufficient for membership in the category so defined. In the classical view, subconcepts can be nested within concepts.

2. The probabilistic view also sees concepts as summary descriptions, but they do not rely upon necessary and sufficient conditions. Instead, probabilists posit that category membership is decided on the basis of some measure of central tendency of an object's features, dimensions, or holistic properties. Nesting of subconcepts is possible using conditional probabilities.

3. The exemplar view does not define concepts as summary representations. Rather, concepts are characterized by descriptions of some of their exemplars or prototypes.

Smith and Medin opt for a mixed approach, one which combines the probabilistic and exemplar views. Although they marshall a great deal of evidence suggesting that people, particularly children, maintain exemplars of concepts, the exemplar view cannot stand alone as a theory of classification. While necessary, exemplars are insufficient. Smith
and Medin [1981: 159-160] find the exemplar approach wanting in two ways.

1. It does not specify principled constraints on the relation between exemplars that can be joined in a concept. That is, it provides no principles for joining two or more possibly disparate percepts together under a concept.

2. It does not account for the acquisition of summary information. An adult may say to a child "all birds lay eggs", but the exemplar view provides no means for representing this summary information.

To remedy these deficiencies, Smith and Medin argue that the exemplar view must be supplemented with one of the views that provides a summary representation. They choose to mix a probabilistic summary with the exemplar view, rejecting the possibility of mixing the exemplar view with a classical-view summary.

Two considerations motivate Smith and Medin to mute the force with which they present their choice. First, Keil's [1979] predictability theory (see section 7.6.2) suggests to them that they may have rejected the classical view too hastily. Second, advocates of neither the probabilistic nor exemplar approaches have provided a satisfactory account of conceptual combination i.e., combining the concepts of "pet" and "fish" to form "pet-fish").

In opting for a mixture of the probabilistic and exemplar views, I believe Smith and Medin have made the wrong choice. A mixed representation of the classical and exemplar views is, in my view, more appropriate and more defensible. Apart from Keil's predictability theory and the problem of conceptual combination, the psychological
evidence reviewed by Smith and Medin fails to cast any doubt upon the classical view as formulated contemporarily.

Before reviewing the Smith's and Medin's interpretations of this evidence it will be useful (a) to discuss the appropriate limits of probability in knowledge representation, and (b) to characterize the classical view as it is understood in contemporary approaches to categorization and knowledge representation in AI.

5.3 The Appropriate Role of Probability

The approach adopted in Relatus to the evaluation and representation of knowledge, like that of its direct antecedent [Minsky, 1974; Winston, 1982] is deterministic, in one sense of that term. Except for knowledge that is explicitly cast probabilistically, probabilistic sorts of inference are avoided, including Bayesian statistical techniques and the fuzzy-set techniques that Smith and Medin [1981: 180-182] advocate.

I do not wish to imply that these techniques are not useful for some purposes. Quite the contrary, they can be and are applied to a wide range of non-human, asocial problems [Henrion, 1987]. But these essentially normative techniques often have little, if any, empirical adequacy.43 Unless so trained, and, even then, only when a particular domain is represented probabilistically, people are simply not probabilistic reasoners [Kahneman, Slovic, and Tversky, 1982].

43 These techniques are normative in that they presume an ethic of ability maximization which is often at odds with Kantian-based fairness ethics as well as the less universalizable yet no less real, particularistic or heteronomous ethics.
Perhaps because so many theories, particularly in economics, presuppose that people do indeed reason probabilistically, these findings have motivated proposals for innovative variations on probability theory. Each proposal attempts to modify or restrict the rules of probabilistic inference in some way so that analytic outcomes more closely resemble the systematically biased (from utilitarian norms) outcomes of human reasoning.

In modeling political cognitions, another alternative is available. One may simply reject probabilistic techniques at the representational level in favor of a deterministic approach, that represents entities as simply present or not present. On this view, representational entities take a probability of one or a probability of zero. No intermediate gradations are allowed.

To be sure, one might represent the various modalities e.g., "That might be a dog"), gradations (e.g., "The man is very bald"), etc., that appear in texts. However, these qualifications may be represented deterministically, as in Relatus, by predicking lexical items with markers indicating modal, adjectival, adverbial, and ordinal qualifications.

Smith and Medin [1981: 179-181] want to ground their notion of fuzzy probability in Zadeh's theory of fuzzy sets, but Zadeh himself rejects this possibility.

... the theory of fuzzy sets is aimed at the development of a body of concepts and techniques for dealing with the sources of uncertainty or imprecision which are nonstatistical in nature. For example, the proposition "x is a small number" in which small number is a label of a fuzzy subset of nonnegative integers, defines the possibility distribution rather than the probability distribution of x.
What this implies is that if the degree to which an integer \( n \)
fits one's subjective perception of \textit{small number} is \( m \), then
\( p \), the possibility that \( X \) may take \( n \) as its value, is
numerically equal to \( m \). Thus, the proposition "\( X \) is a small
number," like the proposition "\( X \) is a number smaller than 5,"
contains no information concerning the probability of the
distribution of values of \( X \). In this sense, the uncertainty
associated with the proposition "\( X \) is a small number" is
nonstatistical in nature [Zadeh, 1980: 421].

As detailed elsewhere [Mallery and Duffy, 1986], the Relatus
implementation is based upon an eidetic, or phenomenal, representation
of literal textual contents. Once an eidetic representation of textual
contents is achieved, later analyses involve successive markings of
lexical tokens represented at the eidetic level with semantic relations
(e.g., relations of synonymy and paraphrase) to other texts that are
also represented lexically.

I certainly do not wish to deny the possibility that Zadeh
fuzziness factors may prove to be analytically useful markings of
lexical tokens for understanding the linguistic hedges in propositions
like "\( X \) is a small number" or "Elmer is very, very tall". I wish only
to deny that fuzziness is are particularly useful for categorization
and conceptual representation. Even if one were to incorporate
fuzziness factors in this way, it would be fundamentally mistaken, for
purely pragmatic purposes, to encode only such fuzziness factors at the
representational level. Such a strategy would force all later
interpretations to be possibilistic, and it would enforce one set of
fuzziness factors upon all later interpretations.

Apart from processes that reason over textual contents that are
explicitly probabilistic (or explicitly fuzzy), explicit numerical
indices of probability or fuzziness associated with representations of
textual contents are neither necessary nor desirable. This is particularly true with respect to categorization, or the representation of relationships between classes and the subclasses and individuals they subsume.

5.4 The Modern View of the Classical View

The version of the classical view critiqued by Smith and Medin is less than fully relevant to contemporary knowledge representation practice in artificial intelligence. In this section, I want to distinguish the classical view Smith and Medin reject the classical view as it is formulated and implemented in contemporary, frame-based, AI knowledge representation systems. In passing, I will summarize Rey's [1983] important objection to the Smith and Medin critique.

5.4.1 The Classical Version of the Classical View

The version of the classical view critiqued by Smith and Medin is essentially the view expounded by Aristotle [Ross, 1927] in the Categories [1a21–2b21; 3b10–4a21] and the Metaphysics [1028a10–1030b6]. On Aristotle's model, a concept is defined by the necessary and sufficient conditions (or defining conditions) a percept must satisfy if it is to be categorized legitimately as an instance of the concept. For example, in order to be categorized as a square, a percept must meet the following conditions.

44 It should be noted that the discussion of the classical view revolves around Aristotle's use of necessary and sufficient conditions. The contents of the higher-level categories he posits are not at all relevant. cf. [Thompson, 1967].
1. It must be a two-dimensional, closed figure.
2. It must have four sides.
3. The sides must be equal in length.
4. All interior angles must be equal in degree.

Each feature is necessary for the object to be classed a square, and these features jointly suffice for squareness.

In his famous analysis of games, Wittgenstein [1953: paras. 66-71] argued convincingly that no set of necessary and sufficient conditions for gamehood can be generated. This critique is widely believed to devastate the classical view of concepts. It appears, then, that only a priori concepts, such as those of mathematics, can be characterized by necessary and sufficient features, while natural kind terms, like "game", cannot. 45

5.4.2 Rey's Objection

Rey [1983], in his critique of Smith and Medin, points out that one must distinguish between the metaphysical (i.e., ontological) and epistemological usages of the classical view. Necessary and sufficient conditions may exist for particular concepts, but we may not know what these conditions are.

For example, supposing the validity of the chemical theory upon which it is based, one can state that "having the atomic number 79", is a necessary and sufficient condition for a substance to be categorized as gold. It would be silly to suggest that one must know this defining

45 Wittgenstein perhaps receives too much credit for this. Kant [1781: 586-588] makes a very similar observation about empirical definition.
condition in order to categorize a substance as gold. But this apparently is suggested by Smith and Medin. They characterize the metaphysical classical view as an epistemological theory, or a theory about how one identifies category membership.

In their response, Smith, Medin, and Rips [1984] grant Rey's distinction, admitting that their considerations affect only epistemological usage of the classical view. Rey, however, went further, suggesting the possibility that classical-view representations of the necessary and sufficient conditions of category membership might be psychologically real, provisionally held stereotypes, which people exploit when making category judgments. Smith, Medin, and Rips reply.

We think this is an interesting suggestion about the naïve or common-sense metaphysics that we may all have as part of our knowledge about concepts . . . , but we are hard pressed to come up with other kinds of psychological implications of Rey's metaphysical version of the Classical view [Smith, Medin, and Rips, 1984: 273].

To me, this psychological implication concerning stereotypes is, by itself, important enough to warrant caution before dismissing claims to the psychological reality of the classical view. Rey's suggestion is full of merit, and I will argue for the psychological reality of an AI version of the classical view. I will make my argument by reinterpreting the experimental evidence adduced by Smith and Medin [1981] to dispute the classical view. The evidence is reinterpreted in light of an AI approach that represents concepts not merely by the necessary and sufficient conditions for category memberships, but also by exceptions to those conditions.
5.4.3 The Contemporary AI View of the Classical View

Suppose for the moment that people actually do adopt, albeit provisionally, necessary and sufficient conditions for category membership. Because, as noted above, these conditions cannot always be expressed for natural kind terms, people must have some representational means for handling the inevitable exceptions.

We may restate this need by reference to Hume's [1739: 1.3.6] problem, which demonstrates the fallibility of all inductive generalization. The problem has two aspects.

1. By what authority can we ever presume that the conjunction of two or more experiences implies that these experiences will always appear so conjoined? Clearly, to claim as authority the generalization that "things in experience often appear jointly in a similar way" will not suffice, since it would ensnare us in an infinite regress. Thus, there can be no empirical authority for inductive generalization.

2. Inductive generalizations lack the apodictic certainty of deductive inferences. As Marvin Minsky once conjectured [personal communication], all rules have exceptions. Even if we have never encountered an exception to a rule like "if x is a swan then x is white", we can never be certain that all swans are white. We surely haven't encountered all the swans that ever existed, much less all the swans that will exist in the future.

Even Minsky's conjecture is a rule that from itself has exceptions. Formal rules, like those of mathematics and logic, must be excepted. For example, "all base 10 whole numbers ending in an even digit are even" is a rule that has no exception, as are the rules of inference in the predicate calculus and the rule about squares noted above. Mathematical, logical rules are not rules about anything existing in experience. They are a priori, posited prior to experience. Only when the concept in question is a natural kind term
-- a concept induced from phenomenal experience -- do we require exceptions.

In terms of the design for knowledge representation systems, the fallibility of induction necessitates a method for handling exceptions to general rules. Probabilistic systems have such a method inherently. The exceptions to any rule in which the consequent follows with $x$ probability from the condition are captured in the complement of $x$. That is, the probability of an exception to a rule is simply $1-x$. For deterministic systems, Minsky [1980] proposes "censors", or *oeteris paribus* or "unless" clauses. On this view, rules have the form, "if $x$, then $y$, unless $z$".

To use Wittgenstein's game example, suppose we provisionally posit the necessary and sufficient conditions of the concept "game" as follows:

1. A game involves two or more players.
2. A game has winning as its purpose.
3. A game is played on a board, table, or playing field.

We can then add "censors" to this set of conditions, in order to handle the obvious exceptions. As examples, we must except solitaire from condition 1, we must except the children's game ring-around-a-rossey from condition 2, and we must except certain "mind games" from condition 3.

The view of the classical view Smith and Medin critique, it should be emphasized, is not this contemporary AI view. Rather, they critique
the version of the classical view articulated in antiquity that includes no provision for exception handling.

While no version of the classical view by itself suffices for a theory of conceptualization and categorization, Smith and Medin observe correctly that the exemplar view by itself also does not suffice. Some form of summary representation is needed. Aside from the reasons Smith and Medin present (noted above), summary representations are required because, in categorization tasks, exhaustive comparisons of phenomenal objects with prototypical representations would be intractable. Moreover, the exemplar view provides no method for handling exceptions. On the other hand, as I've argued above, both the contemporary classical view and the probabilistic view each provide means for handling exceptions. The classical view handles exceptions by admitting ceteris paribus conditions, or Minsky censors, while the probabilistic view would handle exceptions by complementing probabilities.

The real choice, then, boils down to one between (a) a mixture of the probabilistic and exemplar views and (b) a mixture of the classical and exemplar views. While Smith and Medin opt for the former, I disagree and opt, at least provisionally, for the latter.

Before proceeding to a consideration of the arguments Smith and Medin direct against the classical view, an important ancillary point must be clarified. One may certainly encode in concept representations the number of particulars subsumed by that concept. As discussed by Mallery [1986], this information can support the efficient reference of lexical configurations in a semantic network. In such a scheme,
explicit fuzziness factors are not necessary and need not be represented explicitly. They may be readily computed as a function of the number of particulars of a concept that do and do not include a particular defining feature.

I do not wish to argue that there is no fuzziness or uncertainty associated with our understanding of the relationship between concepts and the percepts and sub-concepts they subsume. I argue instead that the probabilities (or fuzziness factors) of subsumption relationships are insufficient for the representation of natural kinds. A conceptual representation that incorporates the defining features (with exceptions) necessary and sufficient to warrant the subsumption of a percept under a concept is far more useful inferentially than a conceptual representation that incorporates only a probability or fuzziness factor for subsumption relations. Simply stated, a penguin is a bird that, among other things, doesn't fly. A penguin is decidedly not a 0.7 bird.

I am now prepared to consider the reasons Smith and Medin reject the classical view. I hope to show that each of their criticisms can be interpreted as being either irrelevant to or supportive of the classical view as conceived contemporarily.

5.5. Smith and Medin on the Classical View

Smith and Medin review both a priori, arguments against the classical view and experimental results that, they believe, militate against the psychological reality of the classical view. I will discuss each in turn.
5.5.1 A Priori Criticisms

Smith and Medin first review four general, a priori criticisms of the classical view, arguing that each fails to devastate it. First the classical view excludes functional features. Smith and Medin point out that this objection can be met merely by allowing functional features into list of necessary and sufficient features defining a concept.

Second the classical view excludes disjunctive concepts. By "disjunctive concepts", Smith and Medin apparently refer to concepts, disallowed by Aristotle in the Metaphysics [1037b8-1039a8], that would subdivide a concept along some orthogonal conceptual dimension. For example, Aristotle would subcategorize quadrupeds into their cloven and non-cloven variants and not into their featherless and feathered variants. To do the latter would result in a category scheme in which there would be as many concepts as there are percepts. Brachman and Levesque [1984] rediscovered this point recently. They observed that such orthogonal disjunctive subcategorizations lead to intractability in the determination of class subsumption relations. Thus, on any view of conceptual structures, orthogonal disjunctive concepts must be disallowed. For their part, Smith and Medin miss the point, suggesting only that disjunctive concepts are rare. In fact, however, in the classical view, non-orthogonal disjunctive concepts are not only included, they pervade.

Third, there are cases for which the appropriate classical representation is unclear. For example, is a tomato a fruit? Most people answer this incorrectly. But, as Smith and Medin correctly observe, this only indicates that many people have simply misidentified
some taxonomic relations. Of course, we can only state that the
taxonomic relations have been "misidentified" if we are prepared to buy
into the botanist's taxonomy.

Fourth there are cases for which the classical view cannot specify
defining features. The canonical example is Wittgenstein's [1953:
paras. 66-71] analysis of the concept of "game", recounted above.
Wittgenstein argued that "game" cannot be described by any set of
necessary and sufficient features, but only by some "family
resemblance". However, Smith and Medin observe, this may only imply
that people differentially make progress toward the complete
representation of the necessary and sufficient features which define
particular concepts.

To me, the latter two objections both presume that individuals
must agree upon ascriptions of category membership. This is a
supposition of the metaphysical usage of the classical view, but it is
certainly not presupposed in its epistemological usage.

Also, we need not understand Wittgenstein's point about family
resemblances as contradicting the classical view. In fact,
Wittgenstein's speculations concerning family resemblances ground the
exemplar approach [Wittgenstein, 1953: para. 75] with which I have
recommended mixing in the classical view. Smith's and Medin's
recognition of the possibility that our inability to specify necessary
and specific conditions for many concepts may only reflect the
incompleteness of our theories about those concepts presages Rey's
suggestion, reviewed above, that an epistemological usage of the
classical view would constitute a theory of stereotypy.
5.5.2 Experimental Criticisms of the Classical View

Smith and Medin counter the a priori objections to the classical view in order to refute the common presumption that these logical objections completely destroy it, obviating any and all psychological inquiry. Smith and Medin then review the more important objections that have arisen from experimental studies of categorization.

5.5.2.1 Typicality Judgments

Smith and Medin first review studies that indicate that respondent judgments of the typicality of an object, or the degree of its "family resemblance" to a concept, vary with experimental manipulations of the distribution of features. Some of the features that experimental subjects use to discriminate the typicality of a particular percept are not common to all members of the category. Thus, contra the classical view, Smith and Medin argue that an account of typicality variations cannot be provided without reference to non-defining features.

Smith and Medin observe that this result fails to devastate the classical view. One might argue, for example, that more common features require more processing time to evaluate than do less common features. This would salvage the classical view, but only by positing ad hoc processing assumptions, which, Smith and Medin assert, would speak poorly for the classical view. But does it?

Barsalou and Bower [1984] report the following. Subjects were asked to learn clusters of medical symptoms associated with various imaginary diseases. Once they were able to name diseases when given a list of symptoms, they were unexpectedly tested in the reverse
direction. They were asked to generate a list of symptoms when given
the name of a disease.

The basic finding was that a symptom's discriminativeness was
strongly correlated with the likelihood that subjects could
recall it from the disease name. If the disease had a
symptom that did not occur for any other disease, subjects
invariably learned that symptom first and later recalled it
better than any of that disease's other symptoms. If a
disease had no unique symptoms, subjects still learned its
most discriminative symptoms. Subjects appeared to learn
distinctive symptoms best because these were most useful in
identifying the disease. [Barsalou and Bower, 1984: 6-7].

Smith and Medin conflate the core defining features of a concept
with procedures for identifying a perceptual instance of the concept.
Distinctive features may simply not be among those that arenecessary
for class membership. For example, although some birds do not fly, it
is rather distinctive of birds that many of them do fly, and one may
reasonably use flying ability as one among many indicators that a
particular percept is an instance of a bird.

The observation that typicality judgments depend upon features
that are not common to all members of a concept is thus explicable as
an artifact of conditions of search in the experimental task, and not
on any deficiency of the classical view of categorization. The
subjects may have chosen to avoid the common features because these
features are less discriminatory, not because they do not enter into a
summary representation of the concept.

5.5.2.2 Nonnecessary Features

Smith and Medin next review experiments that directly indicate
that people utilize non-necessary features when making category judgments. They summarize these results in syllogistic form.

1. Some features listed for a concept were nonnecessary ones (for example, "flies" for bird).

2. These nonnecessary features were correlated with categorization performance.

3. Therefore, nonnecessary features are used in categorization.

Again, the discriminativeness of certain non-necessary features, like "flying" may account for their usefulness in category judgments, thereby obviating the objection. But the objection also falls as soon as one admits ceteris paribus conditions, or censors, to one's classically-oriented theory of concepts. The rule for bird may simply state that (among other things, about features like beaks and feathers) "if x is a bird, then x flies, unless x is a penguin, ostrich, kiwi, dodo, et cetera. Nonnecessary features may certainly be used in classification. In fact, precisely because all rules about natural phenomena admit exceptions, nonnecessary features -- those that discriminate the exceptions -- must be used in categorization.

5.5.2.3 Nested Concepts

Smith and Medin review experiments that bear upon the classical view's assumptions about nested categories. Nested categories refer to concepts that are subsumed by other, more general, concepts, which in turn may be subsumed by still other, even more general, concepts. Features indicative of all particulars subsumed directly by a concept, or indirectly through subsumed concepts which subsume the particular, may be inherited downward in the taxonomic hierarchy.
For example, a particular, phenomenal robin is subsumed by the concept of "robin", which is subsumed by the concept of "bird", which is subsumed by the concept of "animal", which is subsumed by the concept of "thing". The feature "is alive" might appear as part of the summary description of "animal". This feature would be inherited by "bird", by "robin", and by the particular robin. In this way, all nodes subsumed by "animal" in the taxonomic hierarchy need not explicitly represent the feature (unless they are exceptions, like stuffed animals and cartoon animals).

Smith and Medin review evidence indicating that, as the classical view predicts, subjects generally rate immediate superordinate concepts more similar to a concept than they do more distant superordinates. Exceptions to this result have been found however, but Smith and Medin believe that these may be due to the use of relatively unfamiliar concepts, like "alloy" and "mammal". They argue that other exceptions cannot be explained away by familiarity, citing the example of "chicken" and "duck", which are consistently judged to be more similar to "animal" than to "bird".

This exception can easily be "explained away" by another factor, however -- relevance to human purposes. Chickens and ducks, unlike robins, penguins, and many other birds, are simply more acceptable fare for an evening meal. Another explanation, perhaps more likely, is that our naive personal theories of chickens and ducks tend to be walking and swimming things, respectively, not flying things, and birds are flying things, usually.
The important point is that the classical view of categorization, in its epistemological usage, presumes only that subjects will adopt a taxonomic structure of nested concepts. Plausibly, the taxonomic structures any subject adopts will be most well-suited to his or her purposes and experiences. The classical view does not presume that subjects will adopt the taxonomy suited to the purposes of an ornithological specialist. To use Rey's terminology, Smith and Medin the metaphysical (i.e., ontological) usage of the classical view to dispute its epistemological usage. To posit the metaphysical usage as a psychological construct is to deny that interpretations can differ across interpreters.

Smith and Medin pose the use of nested triples as another problem for the classical view. They assert that the classical view predicts that a probe concept should be categorized faster when the target concept is a distant superordinate than when the target is an immediate superordinate.

For example, robin should be categorized faster as an animal than as a bird. But this prediction has often been disconfirmed. The reasoning behind the critical prediction is as follows. For a nested triple, the distant superordinate must contain fewer features than the immediate one -- for example, animal has fewer features than bird -- which is just the third assumption in the classical view at work again. And the fewer features there are in the target concept, the fewer must be compared in the comparison stage of our classical-view model, and the less time is needed to decide that the probe concept is indeed a member of the target. This prediction falls out of any classical-view model that assumes categorization is based on a limited-capacity comparison of probe and target features. [Smith and Medin, 1979: 48].
First, the preoccupation with timing results here constitutes an ad hoc processing assumption which the authors decried only a few pages earlier. Timing studies presume that the processing is performed serially. Second, the experimental results reviewed by Smith and Medin are inconclusive on this point. Third, even if we accept the processing assumption and ignore results which Smith and Medin claim support the classical view, the prediction of the classical view is actually the reverse.

To illustrate, suppose that a subject is asked "Is a robin an animal?" and "Is a robin a bird?" The classical approach (with the additional baggage of the serial processing assumption) would actually predict that the latter would be processed faster. Smith and Medin assume that respondents compare features, so that, in response to the first question, subjects compare the features of animal with those of robin, while, in response to the second question, they compare the features of bird to those of robin. The first comparison includes fewer features, so it should terminate more quickly. However, on the classical view, subjects already maintain taxonomic subsumption links between animal and bird and between bird and robin. No feature comparison is needed to answer the question! Subjects can simply search up the hierarchy of taxonomic links to answer each question successfully.

The search is likely to proceed from lower to higher levels in the taxonomy due to the expansion factor of the hierarchy. The time complexity of search down the taxonomy is an exponential function in the branching factor of the hierarchy, while the time complexity of
search up the hierarchy is a logarithmic function in the branching factor. Thus, the top-down search presupposed by Smith and Medin is combinatorially explosive.\textsuperscript{46} The notion that the classical view, or any other view that posits nested concepts, requires inefficient top-down search in such categorization tasks is simply absurd.

5.5.3 Assessment

In sum, where Smith and Medin have not mischaracterized the classical view, their objections apply only to its ancient formulations. They simply do not affect the contemporary AI formulation.

I believe I have now shown that Smith and Medin have rejected the classical view far too hastily. In conjunction with their admission that the mixture of the probabilistic and exemplar views raises problems that, presently at least, appear irresolvable, one may reasonably conclude that a mixture of the classical and exemplar views remains plausible as a candidate theory of categorization.

This discussion has argued negatively against the reasons Smith and Medin provide for rejecting the classical view. In order to accept the classical view, in its modern formulation, a positive argument for

\textsuperscript{46}In a parallel processing environment, the explosion would be in the number of processors required to perform the task. Etherington and Reiter [1983] report that allowing exceptions (\textit{ceteris paribus} conditions) in such inheritance hierarchies places stringent limitations on the degree of parallelism possible. In general, when consistent inferences are desired, the degree of parallelism is bounded by the number of exception links in the network. This limitation indicates that much processing in inheritance hierarchies is performed in serial fashion. Studies that measure the latency of responses to categorization and synonym detection tasks (reviewed in [Herrmann, Papperman, and Armstrong, 1978]) conform to this view.
its superiority must be presented. In the next chapter, I show how the process of hypothesis formation, including the formation of novel hypotheses, can be explained within the classical framework. Because no explanation of hypothesis formation is forthcoming from advocates of the probabilistic view of concept formation, and because I doubt that such an explanation is possible, this constitutes one indication of the superiority of the classical approach.

A positive, transcendental argument for (the modern version of) the classical view may also be constructed. Before concluding this chapter, let us consider the broad outlines of such an argument.

5.6 A Transcendental Argument for the Classical View

As noted above, the ancient version of the classical view appears to work best, and without the need for ceteris paribus conditions, when a priori concepts, like those of mathematics, are concerned. Only when we move into the domain of natural kind terms does the apodictic certainty of the classical view (in its epistemological usage) break down. At that point, we are forced to admit exceptions to the necessary and sufficient conditions that define concepts.

Although it certainly may become more elaborated, the basic mechanism of conceptualization is itself a priori, or logically prior to all experience. Without such an innate mechanism, no learning would be possible. Putnam [1981: 113-119], who has most clearly articulated the epistemological usage of the classical view as a theory of stereotypy [1975], advances such a transcendental characterization of conceptualization in order to refute Feyerabend's [1970, 1975] methodological anarchism.
The evidence (of sorts) that *a priori* ideas can be characterized by necessary and sufficient conditions constitutes support for the classical view. The story of conceptualization might proceed as follows:

1. We are born with an innate capacity to represent generals (concepts) as necessary and sufficient conditions, and the ability to induce generals from experience.

2. We notice that our inductions of natural kinds tend to be fallible (*pace* Hume, we know them to be necessarily fallible).

3. Rather than giving up induction altogether, we choose instead to mark exceptions to our general concepts or stereotypes.

4. We remain willing to forego our stereotypes in favor of better (i.e., more parsimonious, more coherent, more efficacious, or more consensually validated) formulations.

Language, it seems, plays an essential role in reproducing this knowledge. Levi-Strauss stated it this way:

... [T]o put it precisely, classificatory systems belong to the levels of language: they are codes which, however well or badly made, aim always to make sense [Levi-Strauss, 1962: 228].

In the language we encode our concepts, freezing them in word-capsules that allow us to transmit whatever sense we have made of the world to succeeding generations, however imperfect the sense we have both inherited and remade may be. Smith and Medin appear to demand perfection in this reproductive process. They demand too much from us.

5.7 Our Exceptional Ability

Our profound inability to order the entire range of the world's complexities into an exceptionless framework forces us to add *ceteris*
paribus conditions to the general statements that constitute the intentions or internal relations that characterize our conceptual generalizations. Our ability to except information that might otherwise radically our beliefs is simultaneously both a blessing and a curse. It is a blessing because it affords us the opportunity to maintain a relatively stable stock of basic beliefs upon which we can draw when making the thousands upon thousands of inferences we need to make daily in order to survive and prosper. It is a curse because it can blind us to general truths the evidence for which may like just beyond our noses.

In scientific practice, exceptions allow us to retain useful theories even in light of disconfirming evidence. As Lakatos [1976] makes plain, there are no critical experiments. No single experimental result can falsify a theory because staunch defenders of the theory need only consider the result to be an exception. Only under the weight of very many falsifications would the staunchest defenders feel pressed to reformulate their theoretical constructs and conceptual contents. Precisely how many is uncertain, and Lakatos suggests that most often a "falsified" theory disappears not because of the conceptual reformulations of its adherents, but because of the demise of it adherents.

Counterexamples possess great force in apodictic disciplines, such as mathematics. Goldbach's conjecture, for example, that every even number is the sum of two primes, can be falsified by just one counterexample. One example, however, cannot prove the conjecture
simply because there might be a counterexample. Any proof of the 
conjecture must prove that there can be no counter-examined.

Matters are quite different in non-apodictic disciplines, the 
subject matter of which concerns not \textit{a priori} concepts but concepts 
induced from empirical experience (along with the deductive force of 
any presumed auxiliary theories). Counterexamples there have little 
force because the theory can be excepted incrementally to handle them 
as they arise. In order to convince all but the staunchest defenders 
of the theory, one must provide more than counterexamples. One must 
provide a more satisfactory, more appealing, more useful, or more 
coherent theory.

For example, consider the theory of language acquisition discussed 
in the previous chapter. Suppose we discovered that in some remote 

island in the Pacific there resided a community whose language failed 
to incorporate a condition postulated in Chomsky's theory of Universal 
Grammar. The theory can be defended by simply adding an exception to 
the postulated condition. Thus, it is hopelessly fruitless to critique 
this theory on its own terms. It, like all empirical theories, is 
non-falsifiable.

A more fruitful tact, it seems, would be to offer an alternative 
theory that coheres more strongly the rest of our knowledge, that is 
simpler in the sense that corresponds more directly to reality, that is 
more useful in the sense that it open rather than closes avenues of 
research.\footnote{Chomsky in fact \textit{demands} alternative theoretical formulations 
from his critics and justly so.}
Schlesinger's alternative demonstrates that it indeed has these advantages. Staunch defenders of syntactic exceptionalism may well disagree, but this is to be expected.

Our ability to resist alteration of our beliefs applies within a wide range of social practices, not just in scientific practice. Perhaps the most striking and most famous example is found in the studies by Festinger et al. [1956] of a group that prophesied the imminent arrival of visitors from outer space. When the prophesied event failed to materialize at the appointed time, many group members tend to resist the implication that the prophecy was false. Instead, they continued to search for confirmation, offered new predictions, and redoubled their efforts to proselytize non-believers. In general, the disconfirmation of a highly salient belief creates a "cognitive dissonance" that must be resolve. The resolution strategy is most often not rejection of the salient belief, but rejection of the disconfirmation.

Resistance is a fundamental concept in psychoanalysis, where it is defined as an ego-defense mechanism that prevents patients from producing unconsciously-derived material [Shaffer and Galinsky, 1974: 54-55]. Likewise, it is well known from research on attitude change that people generally resist efforts to persuade them especially when the contents of persuasive communications run counter to the norms of groups to which the individual belongs and for which membership is important to that individual [Hovland, et al., 1953].

These phenomena, dissonance reduction through fact rejection, ego-defensive resistance in the psychoanalytic situation, and resistance to
counter-norm persuasive communications, appear to be deeper than the phenomenon of belief exception relevant to the present discussion. Excepting the intentional contents or internal relations of concepts is something we all must do in order to maintain relative stability in the stereotypes that we necessarily deploy in order to interpret objects and events in the world. Stereotypes, on this view, are not mere falsehoods that inhibit an "accurate" understanding of the "objective" world. They are instead provisionally held summary description — sometimes true and sometime not — that disclose the world to us. They are an essential part of hermeneutical interpretation.

What I am describing is the mode of the whole human experience of the world. I call this experience hermeneutical, for the process we are describing is repeated continually throughout our familiar experience. There is always a world already interpreted, already organized in its basic relations; into which experience steps as something new, upsetting what has led our expectations and undergoing reorganization itself in the upheaval. Misunderstanding and strangeness are not the first factors, so that avoiding misunderstanding can be regarded as the specific task of hermeneutics. Just the reverse is the case. Only the support of familiar and the common understanding makes possible the venture into the alien, the lifting up of something out of the alien, and thus the broadening and enrichment of our own experience of the world [Gadamer, 1966: 15].

Stereotypes, then, constitute a fundamental facet of human interpretation. The intentional contents of our concepts act as foundational presuppositions for the acquisition of additional knowledge, much as in presupposing his theory of optics da Vinci was able to make astronomical discoveries. Likewise, the child's discovery that certain noises can be used to refer to particular relationships allow to bootstrap to a level of competence in which noises refer to particular terms in that relationship. "Ball" becomes "Mommy roll
ball." From the concepts of agent, actions, and patient the syntactic concepts of subject, verb, and object are derived. This in turn serves to bootstrap to a level of competence in which relational configurations become one term in another relation. In other words, "Mommy roll ball" plus "roll to me" become "Mommy roll ball to me." In interaction with others the child further elaborates his linguistic competence, adding determiners, complements, and other syntactic concepts. More complex utterances, such "I want Mommy to roll the ball to me," quickly become possible.

Concepts and stereotypes are of course quite relevant to politics, and the contents of concepts across individuals rightfully deserve an important place in any political science. Brown's [1080] politically-relevant applications of Stephenson's Q methodology have gone a fair distance in demonstrating how disparate perceptions of political objects and events can be across individuals. Research that analyzes and compares across individuals the intentional contents of concepts such as "detente," "appeasement," "justice," "liberty," the various political ideologies, etc., can potentially provide insights into the differing reactions to political events across subjects with disparate world-views as reflected in those conceptual contents.

One can embark on such research in many ways, of course. Following Lane [1962], one might provide common-sense reflections on the contents of in-depth interviews with ordinary citizens. Following Converse [1964], on the other hand, one might proceed by analyzing data collected from large survey samples. The former approach suffers the disadvantage of non-replicability, while the latter suffers the
disadvantage of information poverty. The interpretive presuppositions Lane used informing his common-sense reflections cannot be examined systematically, and the survey data that Converse analyzed are surely not informative enough to disclose the full richness of the conceptual contents of the subjects' political stereotypes.

These approaches are in no sense illegitimate or uninformative. However, by supplementing them with computer simulations of the contents of personal political stereotypes and the inferences make using those contents as premises, we should be able to produce analyses that at once enjoy the replicability of survey analyses and the information richness of interview analyses. An important early step, of course, is the construction of a computational environment in which such analyses can be conducted. Any such environment will embody theoretical commitments concerning the nature of stereotypes. This chapter articulates and defends the commitment in Relatus to the contemporary version of the classical view.

5.7 Conclusions

I have argued that Smith and Medin [1981] too hastily reject the epistemological usage of the classical view of categorization. I have shown that the results of psychological experiments they adduce to disparage the ancient formulation of the classical view are quite compatible with the contemporary AI formulation of the classical view. This latter formulation extends the classical view by admitting ceteris paribus conditions, or censors, that handle exceptions to general rules about necessary and sufficient conditions.
The metaphysical usage of the classical view is a theory about the actual defining conditions of concepts. This usage is most applicable to taxonomic development in the early stages of scientific inquiry in particular domains of knowledge. The epistemological usage, however, constitutes a theory about people's stereotypes. Once the epistemological usage is recognized as a theory about stereotypes, one wonders what functional role remains for the prototypes of the exemplar view. We might indeed maintain in memory eidetic representations of the contents of certain phenomenal experiences, but stereotypes, because they are summary representations can perform many tasks much more parsimoniously than can prototypical or exemplar representations.

This creates a dilemma for the exemplar view. If stereotypes can perform the functional, categorization tasks proposed for prototypes, why bother to grant prototypes any role in categorization at all?

My preliminary view requires prototypical representations for perceptual purposes (but not for classificatory purposes). Prototypes, on this view, are templates used to constrain potentially explosive search processes involved in perceiving phenomenal objects. Since prototypes and stereotypical concepts would be linked directly, successful prototypical template matches could link the percept to the stereotype or concept. Prototypical representations would appear in two varieties, depending upon the degree to which a concept (qua stereotype) has developed.

1. First-order prototypes are eidetic representations of phenomenal objects -- objects that have actually been perceived. These prototypes are maintained in the absence of any stereotypical conceptual formulation. That is, they are most useful prior to the development of a well-articulated
representation of the conceptual contents of a class of objects.

2. **Second-order prototypes** are post-conceptual. They are the imaginary extensional expansions of intentional stereotypes. Once the perceiver/conceiver has experienced enough perceptual instances to formulate a stereotype, expanded stereotypes can then be used as prototypical templates for perceptual constraint.

This classification of prototypes, like the view of classical-view concepts as stereotypes, constitutes only a preliminary formulation. Both require empirical tests.
Chapter 6

Hypothesis Formation

6.1 Introduction

Proponents of any view of concepts and categories must account for the formation of hypotheses. However, unless that account incorporates the propositional contents of the necessary, sufficient, and *ceteris paribus* conditions for concept membership, certain types of hypothesis cannot be explained. This chapter presents a typology of hypotheses and argues that the classical view supports novel hypothesis formation, while the probabilistic and exemplar views do not.

This chapter serves another purpose as well. It counters an important objection leveled at projects like Relatus. The British linguist, Geoffrey Sampson [1979; 1980], arguing from a radical empiricism and a Popperian perspective on hypothesis formation, thinks that computer simulations of human thought are impossible in principle. He believes that computers cannot be made to simulate the creativity that characterizes human mental performance.

Demonstration of the possibility of hypothesis formation is thus critical to AI enterprise in general. It certainly would be a useful capacity for Relatus as well. A creative hypothesis formation capacity would be required for any computational simulation of the process of language learning that does not presuppose the existence of highly complex, innate syntactic principles. Further, a creative capacity would make Relatus far more useful as a tool for social-scientific research than it would otherwise be. This chapter, then, shows how Sampson's objection can be overcome.
To an extent, Sampson's objection has already been demonstrated to be false. The BACON program [Bradshaw, Langley, and Simon, 1983] simulated the induction of scientific discoveries, and Lenat [1982; 1983a; 1983b] showed that heuristically guided simulations can induce new (to the program) concepts and theories from raw data.

Because generalization fails to exhaust the range of human creative competence, these simulations only partially counter Sampson's objection. Some human creative acts are not inductive. Instead, they involve the formation of hypotheses from an extant body of theoretical constructs and theory-relevant data. I will counter Sampson by showing how, under the contemporary version of the classical view, hypotheses, both creative and non-creative, are formed.

6.2 An Overview of Hypothesis Formation

In contrast to Sampson's empiricism, most researchers in both linguistics and in artificial intelligence adopt, implicitly or explicitly, relatively rationalist stances. We saw in Chapter 4 that generative grammarians -- by far the largest group of linguists -- posit the existence of innate mental equipment that affords children the opportunity to learn their native tongues. In Chapter 4, I denied only the syntactic nature of this innate equipment, while accepting the notion that innate ideas facilitate language acquisition. Most, if not all, AI approaches to learning also embrace rationalism, if only implicitly. To switch the discussion from modern "innateness" terminology to the "a priori" terminology of earlier centuries, rationalists believe that we necessarily begin our lives with certain a priori ideas, like those of space and time, without which an empirical
experience would not be possible. Reasonable people may certainly differ with regard to the scope of concepts properly characterized as a priori. However, since we cannot acquire any other ideas without the pre-existence of some acquisition or learning device, it too must belong within that range. Artificial intelligence researchers set before themselves the difficult task of devising demonstrably adequate computational model of these innate apparati.

By "adequate", I mean that the models should be "up to spec". They should exhibit the performance characteristics of intelligent, human mental activity. The process of forming hypotheses constitutes a central ability necessary for any intelligence, artificial or otherwise. Below I attempt to motivate a theory of hypothesis formation that encompasses the formation of two sorts of hypothesis:

1. **Ordinary Hypotheses**: This class of hypothesis involves no extraordinary search or reformulation of knowledge structures. It simply involves reasoning about the "obvious". As discussed below, this form of hypothesis formation is generally known as "abductive inference".

2. **Extraordinary, or Novel Hypotheses**: These class of hypotheses do involve extraordinary search and/or reformulation of knowledge structures. They are far less immediately "obvious" than are ordinary hypotheses, involving some amount of creativity on the part of the person forming them. These hypotheses may be further categorized into two subclasses:

   a. **Normal Novel Hypotheses**: This subclass encompasses those hypotheses that require no reformulation of knowledge structures. They require only a greater amount of search within a pre-existing knowledge structure than do ordinary hypotheses.

---

48 For moderns who feel uncomfortable with 18th-century transcendentalism, it should be noted that we need not count the necessity of these ideas, as Kant [1781] did, as evidence of a human spirit. *A priori* ideas need not be pure. They need only be part of our innate substance, embedded in the genetic code.
b. **Revolutionary Novel Hypotheses:** This subclass encompasses those hypotheses that do require reformulation of knowledge structures. By reformulation, I refer to a reconfiguration of the theoretical constructs that characterize a domain.

The distinction between the two subclasses of novel hypothesis intentionally bears some family resemblance to Kuhn's [1962] distinction between normal science and revolutionary science. Revolutionary novel hypotheses can potentially overthrow paradigms, resulting in entirely new theoretical formulations of particular domains of knowledge. Normal novel hypotheses, on the other hand, are creative acts that presuppose and do not threaten existing theoretical formulations.

### 6.3 Ordinary Hypothesis Formation

We each formulate thousands of ordinary hypotheses every day. For example, suppose we believe this simple conditional:

> If it is cold outside and a nearby window is open, Then I will feel a draft.

Suppose further that we know it's cold outside and we feel a draft. We immediately form the (ordinary) hypothesis that a window is open and we check for an open window nearby. If an empirical test of this hypothesis fails, we form alternative hypotheses. Perhaps the heat is off, perhaps there are cracks in the wall of the building we are in, or perhaps the building simply needs to be better insulated. The procedure for forming such ordinary hypotheses is well-known.
6.3.1 Abductive Inference

Ordinary hypothesis formation is precisely what the noted American philosopher Charles S. Peirce identified as *abduction*. Peirce [1955: 150-156] characterized abduction as follows:

The surprising fact, C, occurs. But if A were true, then C would follow as a matter of course. Thus, there is reason to suspect that A is true.

In syllogistic form, then, abductive inference might appear as follows:

1. C.
2. if A then C.
3. therefore A.

Now of course the inference in this little syllogism does not conform to any of the valid inferences of sentential logic. In the universe of true sentences there may be some other implication responsible for the surprising fact, C. For instance, if B then C might be valid. Does this mean that we should discard Peirce's notion of abduction? Of course not. It means only that abductive inference differs from deductive inference.

Abductive inference has become a rather central notion in artificial intelligence, where it travels under the names of "backward chaining" in expert systems research, "backward inference" in theorem provers, and "goal-driven reasoning" in planning research.\(^{49}\) In their recent introductory text, Charniak and McDermott [1985] explicitly recognize abductive inference as the root of these ideas. However,

\(^{49}\)For a discussion of these ideas, see the relevant entries in [Barr, Cohen, and Feigenbaum, 1982].
they fail to recognize Peirce as their progenitor. Given the magnitude of our intellectual debt to Peirce, and given the rather shoddy treatment he received from the American academy in his own lifetime [Knight, 1965: 1-43], we ought now to be particularly careful to recognize him.

Abduction has import not only for AI, but also for any historically oriented social science. Alker [1985] points out that the virtual absence of training in abductive techniques (and in logic in general) represents a serious deficiency of most contemporary social science graduate programs. Elsewhere [Alker, 1982], he notes that abduction -- "the search for mediating terms [that help] to constitute the essential characters of . . . singular terms" -- represents the sort of intentionalist inference that political analysis at its best achieves. It does not merely describe the "what", it explains the "why". The application of AI expertise in abductive techniques can thus prove exceptionally valuable for the expansion of our sociohistorical analytic capabilities. Let us then examine abductive inference more closely.

Peirce postulated a triad of inferences. Some inferences are deductive, some inductive, and some abductive. The distinction can most readily be seen by referring to another simple syllogism.\footnote{Alker first presented this syllogistic explanation of abduction in seminar.}

1. Nixon is a politician.
2. All politicians are crooks.
3. Therefore Nixon is a crook.
When we infer 3 from 1 and 2, we make a deductive inference (specifically, modus ponens). Deductive inferences which conform to valid rules of inference are apodictic. That is, whenever all the premises are true, the conclusion of a valid deductive inference must also be true.

When we infer 2 from 1 and 3, we make an inductive inference or generalization. Of course, induction requires the assessment of many similar cases before generalizing. Before we can infer the generalization that all politicians are crooks, we must be able to point to many individual cases in which a politician is also a crook. Compared to deduction, we can be less certain of inductive inferences, no matter how convinced we are of the veracity of the premises and no matter how many cases we can point to. There may be an exception, and we may not have been perspicacious enough or patient enough to have discovered one.

When we infer 1 from 2 and 3, we make an abductive inference. With abduction we are on even shakier ground than we are with induction. The validity of an abduction depends upon the range and quality of our experiences. First, as noted above, we may have induced a generalization different from 2 that might account for 1. For instance, instead of the generalization that all politicians are crooks, we may have induced the rule that all scientists are crooks. Alternatively, another generalization may exist that we have as yet, because of our limited experience, failed to induce. Also, the particular case in question may represent an exception to the rule 2. (e.g., All politicians are crooks, except Nixon).
If abductive inference is so uncertain, why bother with it at all? The answer is simple. Abductive inferences generate hypotheses. Whenever we observe some surprising fact, we are usually able to generate plausible hypothetical reasons for it. We do so by performing an abductive inference, collecting generalizations any of which, if relevant to the situation, might explain the surprising fact. In order to convince ourselves that the surprising fact is indeed explained by any abduced reason, we put it to deductive test.

For example, suppose we are told that Nixon is a crook and we then employ the generalization that all politicians are crooks in order to abduce the hypothesis that Nixon is a politician. We can test this hypothesis by testing deductively some other generalizations we may have acquired. These would include other possible consequences of Nixon being a crook or other abducible reasons for Nixon to be a politician.

Abductive inference, then, is a technique for explaining a singular proposition ("Nixon is a crook") by applying a general proposition that might be part of a theoretical formulation ("all politicians are crooks") to produce an hypothesis that might explain the singular term ("Nixon is a politician"). To see that this in fact is what political analysts do, consider the singular proposition "group X experiences difficulties convincing potential members to join".

Given this fact, a political scientist might apply a theoretical term from Olson's [1965] theory of groups. The general proposition might then be "potential members of groups with large potential memberships have an incentive to enjoy the benefits of group membership
without paying the costs". From here the competent political analyst would abduce the hypothesis that "group X has a large potential membership". Should an empirical test of this hypothesis demonstrates that this indeed is the case, the political analyst (now donning the cap of the political consultant) might apply another proposition from the theory of groups and counsel the group's leadership to increase the collective benefits of membership (e.g., offer a magazine, discounts, group insurance, etc., to the members).

6.3.2 Abduction and the Classical View

In the preceding chapter, I defended the contemporary version of the classical view of concepts and categories against objections presented by Smith and Medin [1981], who instead advocate a (loosely) probabilistic model. Its ability to support abductive inference constitutes a critical advantage of the classical view over the probabilistic model that Smith and Medin advocate. Without some representation of the contents of the singly necessary and jointly sufficient defining features of categories, as well as the contents of ceteris paribus exceptions to those conditions, abductive inference would not be possible.

To see why this is so, let's return to the Nixon example and suppose that we believe the following:

1. Nixon is a politician.
2. All politicians are crooks, except Republicans.
3. Nixon is not a crook.
This represents something of an anomaly. In order for the deduction of 3 from 1 and 2 to be valid, Nixon must satisfy the *oeteris paribus* condition on 2. We abduce that Nixon is a Republican -- an empirical proposition that question which we may subsequently test deductively.

This much is straightforward. However, when we approach this little knowledge base from a probabilistic or fuzzy reasoning perspective, the ground seems to disappear from beneath us. Suppose the beliefs above were represented as follows.

1. Nixon is a 1.0 politician.
2. 0.9 politicians are crooks.
3. Nixon is 1.0 not a crook.

Here we are faced with an insurmountable obstacle. When we notice that proposition 3 represents an exception to proposition 2, licensed by the 0.9 fuzziness factor on politicians, we don't know what to do other than to chalk up Nixon as an exception to the general rule 2. We are left with nothing to hypothesize! The probabilistic (or fuzzy) approach fails because it does not allow for this sort of abduction.

6.3.3 The Insufficiency of Abduction

Abduction does not suffice. This form of hypothesis formation requires the pre-existence of a theory. But what if our knowledge of related theoretical constructs leads us to abduce hypotheses that fail when tested deductively? How do we acquire hypotheses that the relevant theories would not predict as potential explanations for an anomaly? Clearly, it must be possible to formulate such hypotheses, since theoretical development would otherwise not be possible. So where do novel hypotheses come from?
6.4 Extraordinary Hypothesis Formation

In the discussion below, I will move rather freely between terminology in artificial intelligence and in the theory of knowledge. The underlying assumption is that there is no substantive difference between common-sense reasoning, a concern of AI, and scientific reasoning, a concern of the theory of knowledge. This assumption is based on two convictions.

1. The quality of any scientific work depends upon the degree to which the scientist applies common sense to the endeavor.

2. Knowledge of the processes by which a machine may acquire knowledge is likely to be acquired best through examination of our knowledge of the processes by which we acquire knowledge.

6.4.1 The Fallibility of Induced Hypotheses

Robbie the Robot needs some way of making sense of the world. Unless we are willing to limit the range of Robbie's activities to some extremely constrained domain, we cannot pre-program him with a "script" for every possible contingency and set him off to deduce his way through the world.

Everything simply cannot be innate. Robots designed to perform in a wide variety of settings must learn. They must somehow acquire general knowledge from specific experiences by testing hypotheses about the world. But where do these hypotheses come from?

51 Of course, scientists endeavor to ground their reasoning in generalizations they believe to be exhaustively tested and confirmed [Rescher, 1970: 9-10], but this is a formal, not a substantive, distinction between common sense and scientific sense.
The simplest answer, of course, is that Robbie induces this general knowledge from experience. Here is a flying thing. Someone tells Robbie it is a bird. Robbie remembers that there exists a thing called a bird that flies. Here comes another flying thing. Robbie asks "Is it a bird?" "Yes," comes the reply, "You can feed bread crumbs to birds." Robbie feeds bread crumbs to the bird, and induces the following generalizations:

- All birds can fly,
- All flying things are called birds, and perhaps (if Robbie believes his interlocutor),
- All birds can be fed bread crumbs.

When Robbie next sees another flying thing, he may abduce the hypothesis that it's a bird. Or, when he next encounters something called a bird, he may abduce the hypothesis that this bird can fly.

A familiar problem arises -- one raised most forcefully by Hume [1739: 1.3.6] and one that Robbie might notice when he first encounters a penguin or tries to feed bread crumbs to a Boeing 757. This problem has two aspects.

1. By what authority does Robbie presume that the conjunction of two or more experiences implies that these experiences will always appear so conjoined? Clearly, to claim as authority the generalization that "things in experience often appear jointly in a similar way" will not suffice, since it ensnares us in an infinite regress. We cannot have induced the generalization that things often appear jointly in a similar way without first having induced that same generalization. Thus, there can be no empirical authority for inductive generalization.

2. Inductive generalizations lack the apodictic certainty of deductive inferences, and for this reason generalizations
require censors, or *oeteris paribus* conditions, as discussed in Chapter 5.

Popper [1959] refers to Hume's problem to illustrate the non-existence of the principle of verification for which logical positivists searched. In place of a verification principle, Popper proposed a *falsification* principle. He believes that scientists should conjecture general laws, stating at the same time the experimental conditions under which any such conjecture must be rejected. The scientist then performs these experiments, trying her darnedest to refute the conjecture, submitting both the conjecture and the experimental results to the critical scrutiny of her peers.

Popper does not infer from Hume's problem that inductive generalization should be discarded. Quite the contrary, he recommends that scientists conjecture generalizations *boldly*. Alluding to Kant, Popper [1959: 368n] asserts that a transцendental argument can be made for inductive generalization. Although no empirical evidence can prove an inductive generalization, the capacity to generalize is required for an empirical experience to be possible.

Popper cautions not to take this transcendental argument too far. It does not follow that our capacity to generalize from experience leads to *certain* knowledge. All generalizations should be accepted provisionally, and we should constantly be on our guard for the appearance of exceptions. So should Robbie.

Popper's logic of scientific discovery has of course been modified by his students, who show, for example, that critical experiments have not characterized scientific progress historically [Kuhn, 1962] and
that a sophisticated falsificationism would be more conservative with respect to refutations [Lakatos, 1970]. But what interests us here are not harsh refutations, but rather bold conjectures. How do they arise?

6.4.2 The Popperian View of Novel Hypothesis Formation

Popper believes that there can be no theory of creative hypothesis formation or of the generation of new ideas. Brand new ideas are, in his view, not illogical — they are alogical. Popper [1959: 32] claims that

... there is no such thing as a logical method for having new ideas, or a logical reconstruction of this process. My view may be expressed by saying that every discovery contains "an irrational element", or a "creative intuition", in Bergson's sense. In a similar way Einstein speaks of the "search for those highly universal laws ... from which a picture of the world can be obtained by pure deduction. "There is no logical path", he says, "leading to these ... laws. They can only be reached by intuition, based on something like an intellectual love ("Einfühling") of the objects of experience."

Apart from the fact that this represents a rather mushy-headed account of novel hypothesis formation from a philosopher as hard-nosed as Popper usually is, it represents a critical problem for AI. If indeed the process of novel hypothesis formation admits of no logical characterization, then Robbie is doomed forever to labor within the harsh confines of highly constrained and rigidly pre-specified domains. In Kuhnian terms, he can only practice normal science, never revolutionary science. He can never have a new idea. Drawing upon this Popperian view, Sampson [1980: 99], in fact, argues that the
logical character of novel hypotheses sharply delimits the scope of AI.

Any Artificial Intelligence model, unless it restricts itself to some peripheral area in which these questions [of the source of novel hypotheses] do not arise, must presuppose that . . . unpredictable innovation . . . does not occur.

Minsky [1986] refers to such objections as "old superstitions". He suggests a refutation that succeeds in principle. When confronted by an apparently insoluble dilemma, Robbie may simply generate all possible combinations until he discovers one that works.

To use Minsky's example, suppose Robbie wants to build a bridge across a stream. Suppose further that only some sticks and stones or boards and nails are lying about. Suppose that Robbie has no theory of bridge-building. What shall he do? Robbie randomly manipulates combinations of the items in his perceptual field until one such combination results in a bridge across the stream. This solution, while succeeding in principle as a counter to Sampson's Popperian objection to AI, of course does not succeed in practice. As Minsky observes, it will work so long as we (and Robbie) are willing to wait long enough.

But we're not willing to wait that long. We require a method of generating novel and potentially successful hypotheses in a practicable of time. Even for Robbie's simple stream-crossing problem, the possible combinations that might be generated could consume eons. So, if not at random, how are potentially successful novel hypotheses generated?

4.4.3 Psychological Treatments of Novelty
The question of novelty is by no means new. It perplexed William James, for instance, whose associationalist psychology was firmly rooted in the classical approach to conceptualization and learning [Tatarkiewicz, 1973: 19].

If pure thought runs all our trains, why should she run some so fast and some so slow, some through dull flats and some through gorgeous scenery, some to mountain-heights and jewelled mines, others through dismal swamps and darkness -- and run some off the track altogether, and into the wilderness of lunacy? Why do we spend years straining after a certain scientific or practical problem, but all in vain -- thought refusing to evoke the solution we desire? And why, some day, walking down the street with our attention miles away from that quest, does the answer saunter into our minds as carelessly as if it had never been called for -- suggested, possibly, by the flowers on the bonnet of the lady in front of us, or possibly by nothing that we can discover? If reason can give us relief then, why did she not do so earlier [James, 1890, v. 1: 551-552].

James' answer rests upon his conception of the roles of discrimination and association in individual psychological development. For James, our knowledge advances by alternating sequences of discrimination, in which wholes are analyzed into their component parts, and association, in which objects appearing separately are conjoined to form new wholes [James, 1890, v. 1: 550]. These alternating sequences are common to all people, according to James. He believed that the distinction between the creative genius and the remainder of the population may be attributed to differences in the way the genius associates. Everyone associates by contiguity, according to James. We associate in our imaginations objects that tend to appear

---

52 For a more modern account of this view of conceptual development, see Harvey, Hunt, and Schroder [1961].
together. The creative person, to James, associates not only by contiguity, but also by similarity or analogy.

What does scientific man do who searches for the reason or law embedded in a phenomenon? He deliberately accumulates all the instances he can find which have any analogy to that phenomenon; and, by simultaneously filling his mind with them all, he frequently succeeds in detaching from the collection the peculiarity he was unable to formulate in one alone; even though that one had been preceded in his former experiences with all those which he now at once confronts it ..... The mind in which this mode of association most prevails will, from its better opportunity of extricating characters, be the one most prone to reasoned thinking; whilst, on the other hand, a mind in which we do not detect reasoned thinking will be one in which association by contiguity holds almost exclusive sway [James, 1890, v. 2: 346-348].

In reaction to the 1957 Soviet launch of Sputnik, a cottage industry emerged in the psychological study of creativity. Just as Popper had conceived it, this topic had previously most often been treated as some mysterious phenomenon the understanding of which, by definition, lies beyond the capacities of common man [Razik, 1967]. James' understanding of creativity is reflected in much of this literature. Koestler [1964], for example, writes of a "bisociative" process, in which a situation or idea is perceived in two self-consistent but habitually incompatible frames of reference. The creative act, according to Koestler, always occurs on more than one referential plane. He presents numerous historical anecdotes of creative discovery to back his claims.

For Koestler [1964: 178], the act of creation requires the attainment of a dream-like state of reverie not unlike the state that Kubie [1958; 1967] calls "pre-conscious". From a psychoanalytic perspective, Guilford [1963] refers to pre-consciousness as "ego
expansion" or "undifferentiated spatio-temporal scanning". In this state, controls on mental operations are relaxed. Social taboos, the constraints imposed by language, and the dogmas of common sense, are held in abeyance. For Kubie, the creative individual in the pre-conscious state incessantly assembles and dissembles many diverse patterns. At the decisive moment of discovery, claims Koestler, all the norms of disciplined reason are suspended and the mind is free to drift aimlessly.

The post-Sputnik psychological literature on creativity, in short, seems as fuzzy as Popper's understanding of novel hypothesis formation. 53 I certainly do not deny the remarks of Koestler and Kubie, much less those of Bergson and Einstein. But surely we can characterize creativity more precisely.

Also, must all creative acts be so dreamy? Mental rumination and flights of cognitive fancy might precurse the creative acts that are so extraordinary that they revolutionize the way we understand an entire domain of reality. Can we not also hypothesize creatively within the confines of the related theoretical constructs that partly characterize an existing paradigm?

Feats like the formation of the revolutionary novel hypotheses of Einstein, which result in the entire reformulation of relationships between the theoretical constructs of a domain of inquiry, are not to be expected very often. The formation of such extraordinary hypotheses plausibly requires a great deal of serendipitous search, as suggested

53 For a more thorough review of the post-Sputnik creativity literature, see [Duffy, 1980].
by Minsky's generate-and-test example. It might also plausibly require, in the language of psychoanalysis, the suspension of ego-defensive functions and the abeyance of superego constraints. The dreamy reverie of pre-conscious processes might plausibly facilitate this, although one desires a more precise account.

Creativity, however, is not all that rare. We suspect that creative people live among us, and we have all witnessed incidents in which some creative thinker postulated an hypotheses to which we would ascribe the term "creative". The formation of normal novel hypotheses depends upon the "wit" or "insight" of the hypothesizer. In order to counter Sampson's objection some elaboration is needed of the process by which normal novel hypotheses are formed.

6.4.4 Normal Novel Hypotheses

I conjecture that normal novel hypotheses differ not only in degree, but also in kind, from revolutionary novel hypotheses. They differ in degree to the extent that some relaxation of the social norms reflected in ceteris paribus conditions seems requisite for some novel hypotheses. They differ in kind to the extent that the search for normal novel hypotheses seems far more constrained than the search for revolutionary novel hypotheses. They also differ in kind in that normal novel hypotheses do not involve the reformulation of the relationships between theoretical constructs. Unlike the formation of revolutionary hypotheses, normal hypotheses leave these relationships relatively intact.

Once confronted by a problem that may have perplexed others, the insightful thinker can occasionally utter a creative hypothesis almost
immediately, or after only an astonishingly short period of rumination. This immediacy suggests that normal novel hypotheses, unlike the revolutionary sort, involve a highly constrained degree of search. It also suggests that this search is not random, as in the generate-and-test regime. It suggests some degree of systematicity.

Minsky's [1980] discussion of the relation of jokes to the logic of the cognitive unconscious is particularly relevant for the explication of normal novel hypothesis formation, if only for the simple reason that wit and humor seem to covary across individuals. Extending Freud's theory of jokes, Minsky suggests that censors foreclose certain mental pathways that lead to ludicrous results. Censors, on this view, perform functions beyond merely marking the ceteris paribus conditions of general propositions. They suppress circularities, paradoxes, and other forms of nonsense that might otherwise arise regularly and paralyze thought. They may also represent internalizations of social taboos.

If the violation of censor restrictions apply to the formation of normal novel hypotheses, something like Kubie's "pre-conscious processing" or Guilford's "dedifferentiated spatio-temporal scanning" might facilitate it. On this view, revolutionary novel hypothesis formation simply involves a greater degree of censor censorship than does the formation of normal novel hypotheses. In both, the retraction, hypothetical modification, or censorship of a censor or censors allows the hypothesizer to gain epistemic access to hypotheses that might otherwise have been suppressed.
Revolutionary hypotheses differ in kind from normal hypotheses in that, if validated, they result in the restructuration of the constructs central to an entire domain of knowledge. Validated normal hypotheses, on the other hand, only twiddle constructs at the periphery of some knowledge domain.

The distinction can perhaps be seen more clearly when we consider the residue left by each. Because they radically reformulate entire domains of knowledge, successful revolutionary hypotheses disclose a new world of abducible hypotheses. To express this thought in Lakatosian terms, revolutionary hypotheses make irrelevant, for those who believe such hypotheses to be successful, the protective belt of ceteris paribus conditions that surround a theory. By unravelling the protective belt, successful revolutions make perceptible hypotheses that had been undetectable within the prerevolutionary regime. Normal novel hypotheses, on the other hand, disclose little beyond themselves.

The three classes of hypothesis may be ranked in terms of the degree of search each requires. Ordinary hypotheses (abductions) require little search at all. From a singular statement ("Nixon is a crook") and a theoretical construct ("If X is a politician, then X is a crook") one simply abduces the premise(s) that would make the singular statement comprehensible in terms of the theory ("Nixon is a politician"). In the absence of an abducible hypothesis or when abduced hypotheses fail deductive tests of relevance, an hypothesizer may cast about locally in search of a normal novel hypothesis, relaxing ceteris paribus constraints within a relatively highly constrained search space. If the hypothesizer fails to find such an hypothesis, a
much less constrained search begins. More ceteris paribus constraints may be relaxed, and the hypothesizer may even tentatively construct alternative views of the relationships between theoretical constructs. These alternatives constitute candidate revolutionary theories. The hypotheses abducible from them we have termed "revolutionary novel hypotheses".

The characteristics of search explain how hypotheses are formed in real-time, and they also explain why some hypotheses appear more frequently than others. Because almost no search is needed, simple abductions are formed instantaneously. Revolutionary hypotheses, on the other hand, involve tremendous amounts of search within alternative spaces of conceptual relations. Normal novel hypotheses involve an intermediate amount of search — more than is needed for simple abduction, but, because they involve no hypothetical reformulations of theoretical constructs, far less than is needed for revolutionary hypotheses.54

For an hypothesizer to consider a revolutionary hypothesis plausible, the theoretical reformulation that spawned it should be considered more parsimonious than its forebear. By parsimony I mean that the resulting theory should be more internally and externally coherent than the theory it replaces. It should require fewer ceteris

54 One might object that I am here assuming the ad hoc processing characteristics I decried in the preceding chapter. However, the search for plausible hypotheses undoubtedly exploit the subsumption relations that link percepts to concepts and less abstract concepts to more abstract concepts. However, the assumption of a degree of seriality is warranted on the ground that parallelism is limited in subsumption hierarchies that admit ceteris paribus conditions [Etherington and Reiter, 1983].
paribus conditions and it should provide a better fit to other domains of knowledge, it should be more consistent with empirical experience, it should have more practical utility, and it should be more acceptable by others in the linguistic community concerned with the content of the theory. Since other domains, empirical experience, one's goals, and relevant others are all represented as concepts and percepts, coherence can be considered the overarching desideratum of revolutionary theoretical formulations.\footnote{See Habermas [1973] for a detailed discussion of the various theories of truth to which I allude here. Goodman [1978: 120-125] argues that any account of truth must recognize that ascriptions of truth to any proposition are relative to the version of the world known by the believer. To me, this suggests that coherence, or what Putnam [1983: 65] terms "goodness of fit," can incorporate all these aspects of truth.}

6.4.5 Generative Metaphor

Normal novel hypotheses differ from revolutionary novel hypotheses in another crucial sense. To employ James' terminology, the formation of normal hypotheses exploit association by contiguity. These hypotheses are not, like revolutionary hypotheses, the result of long rumination, suggested by "by the flowers on the bonnet of the lady in front of us", as James put it. Because we are, as Simon [1979; 1983] has argued most forcefully and convincingly, exceedingly limited processors of information, a more systematic and constrained search is normally conducted.

To illustrate how the search for normal novel hypotheses exploits contiguous associations in memory, it will be useful review an example drawn from Schon's [1981] account of "generative metaphor". Schon
argues that the critical aspect of problem solving is problem setting. The way in which a problem is framed often determines whether a solution is found, and always determines the character of that solution. Different people may frame the problem differently. Each exploits metaphorical or analogical skills to "see" the problem in a particular aspect.

As an example, Schon refers to a problem posed to a research group. They were to improve the performance of a paintbrush with synthetic bristles. Compared to a brush with natural bristles, the synthetic-bristle model performed poorly. It "glopped" the paint onto a surface unevenly. The researcher tried a number of improvements, such as altering the bristle diameter and splitting the ends of the bristles. All these efforts proved futile.

Then someone observed, "You know, a paintbrush is a kind of pump!" He pointed out that when a paintbrush is pressed against a surface, paint is forced through the spaces between bristles onto the surface. The paint is made to flow through the "channels" formed by the bristles when the channels are deformed by the bending of the brush. He noted that painters will sometimes vibrate a brush when applying it to a surface, so as to facilitate the flow of paint [Schon, 1981: 257].

The metaphor "a paintbrush is a pump", set the problem in a new light, and the researchers exploited the new problem-setting for all it was worth, leading to a variety of inventions.

According to Schon, the researchers were able to see the paintbrush as a kind of pump long before they could articulate the

---

56 Simon and Hayes [1976] present experimental evidence supporting this assertion. See also [Simon, 1985].
respects in which the two are similar. They first had an intuitive "feel" that the two were similar. Only when they later tried to account for this feeling of similarity were they able to express the similarities. This suggests that the cognitive processing leading to the novel hypothesis that a paintbrush is a pump was in no way conscious.

It would be seriously misleading ... to say that, in making their generative metaphor, the researchers first "noticed certain similarities between paintbrushes and pumps." For the making of generative metaphor involves a developmental process. It has a life cycle. In the earlier stages of the life cycle, one notices or feels that A and B are similar, without being able to say similar with respect to what. Later on, one may be able to describe relations of elements present in a restructured perception of both A and B which account for the pre-analytic detection of similarity between A and B, that is, one can formulate an analogy between A and B. Later still, one may construct a general model for which a redescribed A and a redescribed B can be identified as instances. To read the later model back onto the beginning of the process would be to engage in a kind of historical revisionism [Schon, 1981: 260].

I believe that Schon here stresses the fact that in forming creative hypotheses, the hypothesizer does not consciously search for similarities. Indeed, as discussed above, consciousness of the processing might itself reinforce the censors that inhibit the formation of the hypothesis. But from this one cannot infer that creative hypothesis formation is in principle indescribable or, as Popper would have it, alogical. So let's flirt with historical revisionism and perform a thought experiment regarding the information that the creative researcher-hypothesizer might have pre-consciously processed.
Schon notes that the concepts "paintbrush" and "pump" are both subsumed by the more general concept "tool". Perhaps class relations played an important part in the formation of this hypotheses.

6.4.6 A Simulation of Generative Metaphor

To demonstrate how one might arrive at the intuition of paintbrush-as-pump, the following text was represented in Relatus.

A tool is an object. Objects are things. Substances are things. Liquids are substances. Water is a liquid. Paint is a liquid. Putty is not a liquid. Putty is a substance. A computer is a tool. Computers process information. Information is an abstraction. Abstractions are things. A pump is a tool. A paintbrush is a tool that painters use. Putty-knives are tools that painters use. Paintbrushes apply paint to surfaces. Putty-knives smear putty on surfaces. All pumps pump liquids.

Parsing and referencing this text created a taxonomy under "thing". Some of the objects categorized included necessary and sufficient conditions (no ceteris paribus conditions were included in this toy example), or internal relations. For example, "computers process information" and "all pumps pump liquids" express internal (i.e., essential) relations of computers and pumps, respectively.

A simple program was then written that searched for an hypothesis to explain the behavior of a paintbrush. The central procedure generalized the object of interest (paintbrush at first) and looked at the immediate subclasses of that class to see whether any of them had internal relations whose terms matched the internal relations of paintbrushes. When a match was found, the program suggested an hypothesis. The first hypothesis suggested, as shown in the trace of the program in Figure 6-1, was that paintbrushes smear paint. The
hypothesis was drawn from the knowledge that a putty-knife, another tool that painters use, smears putty and putty, like paint, is a substance. The program asked the user to test the hypothesis, and the user reported that the test failed.

Having exhausted the (known) universe of tools that painters use, the program recursively generated the next level of generalization, the class of tools. It found that computers are tools. However, computers process information and because information and paint only share the most general class, "thin", the program generated no hypothesis. It next found another tool, "pump", and it noticed that pumps pump liquid, which, like paint, is a substance. So it generated the hypothesis that paintbrushes pump paint and invited the user to test the hypothesis. The (pretend) hypothesis test succeeded, so the program invited a knowledge reformulation. First, it asked whether to add the knowledge that paintbrushes pump paint. The user responded affirmatively, so that new internal relation of paintbrush was added. Next it asked whether to subsume paintbrush under the class of pumps. Again, the user responded affirmatively so the class subsumption link was added. Finally, the program asked whether to remove paintbrush from the class of tools that painters use. The user responded negatively, so this subsumption link was retained.
(simulate-hypothesis-generation 'paintbrush)

Finding other nodes subsumed by PAINTBRUSH's classes.
Nodes found under TOOLS THAT PAINTERS USE: PUTTY-KNIVES.

Testing PUTTY-KNIVES.
The internal subject relation of PUTTY-KNIVES is:
PUTTY-KNIVES SMEAR PUTTY.
PAINT is subsumed by LIQUID, SUBSTANCES.
PUTTY is subsumed by SUBSTANCES.
PAINTBRUSHES APPLY PAINT and PUTTY-KNIVES SMEAR PUTTY.
PAINT and PUTTY are both SUBSTANCES.

Test the hypothesis that PAINTBRUSHES SMEAR PAINT.

Did the hypothesis test succeed? (Y or N) No.

Nodes found under TOOLS: TOOLS THAT PAINTERS USE, COMPUTERS, PUMPS.

Testing TOOLS THAT PAINTERS USE.
TOOLS THAT PAINTERS USE subsumes PAINTBRUSHES,
so it cannot be a source for novel hypotheses.

Testing COMPUTERS.
The internal subject relation of COMPUTERS is:
COMPUTERS PROCESS INFORMATION.
PAINT is subsumed by LIQUID, SUBSTANCES.
INFORMATION is subsumed by ABSTRACTIONS.
The internal relations of PAINTBRUSHES AND COMPUTERS do not match.

Testing PUMPS.
The internal subject relation of PUMPS is:
PUMPS PUMP LIQUID.
PAINT is subsumed by LIQUID, SUBSTANCES.
LIQUID is subsumed by SUBSTANCES.
PAINTBRUSHES APPLY PAINT and PUMPS PUMP LIQUID.
PAINT and LIQUID are both SUBSTANCES.

Test the hypothesis that PAINTBRUSHES PUMP PAINT.

Did the hypothesis test succeed (Y or N) Yes.

Considering whether to reformulate knowledge.
Add the relation PAINTBRUSHES PUMP PAINT? (y or N) yes.
New internal relation added: PUMP-2

**Figure 6-1:** Trace of a simulated novel hypothesis formation
(continued on next page)
Should the class of PAINTBRUSHES be subsumed by the class of PUMPS? (Y or N) Yes.

PUMPS now subsume PAINTBRUSHES

Retract PAINTBRUSHES'S membership in the class of TOOLS THAT PAINTERS USE? (Y or N) No.

**Figure 6-1:** Trace of a simulated novel hypothesis formation (cont.)

This simulation could be improved of course by increasing the amount of knowledge represented and by considering the class membership of the verbal relations in the internal relations of the class terms. Nevertheless, it serves to illustrate how novel hypotheses might be mechanically produced from an extant body of knowledge, exploiting the contiguous relationships in the representation of that knowledge.

William James didn't have it quite right. Association by contiguity need not be limited to Peircean abduction. A whole class of creative hypotheses -- by far the largest class in fact -- exploit association by contiguity in their formation. These are the normal novel hypotheses. The contiguous associations in the thought experiment above are all subsumption links, but these may not exhaust the contiguous associations crucial to novel hypothesis formation. Perhaps the Aristotelian aitia -- formal cause, material cause, efficient cause, and final cause -- constitute the vital links.\(^{57}\)

Perhaps any and all links may be involved.

---

\(^{57}\)For a modern discussion of the aitia, see [Moravcsik, 1981].
6.5 Conclusions

I believe I have now characterized the processes of hypothesis formation in enough detail to warrant rejection of Sampson's objection. I certainly have not demonstrated that the account of hypothesis formation presented above is the correct account. Experimental studies and computational simulations may one day provide evidence leading to an elaboration, modification, or outright rejection of this account in favor of some other account. I have shown, however, that Sampson rejects artificial intelligence far too hastily. Grounded solely in the Popperian assumption of the alogical character of creative hypothesis formation, Sampson unfairly attacks a budding, progressive research program.58

Sampson is quite correct in one respect, however. He asserts that innovation is in principle not predictable. Predictions of the creative hypotheses that might arise in any particular mind would require a fairly complete and consistent model of the contents of that mind. At present and for the foreseeable future, our ability to construct such models falls far short. Even if we were able to construct such models, our limited capacity for processing information

58Lakatos [1970: 151–157] argues that a budding research program should be sheltered from powerful, established rivals so long as it would represent a progressive problemshift in the absence of its rival. By a "progressive problemshift", Lakatos means that the research program predicts new phenomena. Artificial intelligence, insofar as it is applied to psychological and epistemological questions, constitutes a progressive problemshift. Simon's [1979: 1983] model of human satisficing, to cite one prominent example, predicts that humans can process only a limited amount of information. This result has important social scientific implications, casting doubt on empirical results and normative theories that assume people have access to complete information regarding particular states of affairs.
would sharply delimit our ability to understand the models and make 
predictions from them. Moreover, if the persons whose creative 
hypotheses we try to predict themselves have access to this modeling 
technology, the task of predicting those hypotheses becomes even more 
fruitless.

The purpose of this chapter has also been served. The view of 
hypothesis formation presented here buttresses the argument in the 
preceding chapter in defense of the modern version of the classical 
view of concept formation. As argued above, even the most ordinary 
hypotheses -- the products of abductive inference -- require more 
information from concepts than fuzzy set membership indices. What is 
true of ordinary hypotheses is true in spades for novel hypotheses. 
Without information pertaining to the contents of concepts like 
"paintbrush" and "pump", like "mass" and "velocity", novel hypotheses, 
whether normal or revolutionary, are literally inconceivable.
Chapter 7

The GEL Shell: A Generic, Extensive Lexicon Shell

7.1 Introduction

Most contemporary linguistic theories and AI natural language implementations are lexically based. They grant to the lexicon a central role [Berwick, 1983: 385]. The movement toward a lexical basis has led linguists to devote more attention to the structure of the lexicon. Likewise, computational lexicons might be developed to provide tools that facilitate both the development and application of AI natural language systems. This chapter reports the progress of one such implementation.

This implementation is not, at this writing, a component of Relatus, although we intend soon to integrate the two systems. At present, Relatus uses a 16,000-token lexicon. This lexicon is implemented as a hash table the keys of which are English words and the values of which are association lists. The association lists store information pertaining to the word's possible categories (parts of speech), its senses, its uses, and its subcategories. Subcategorical information pertains to features of particular words, such as the gender and number of nouns and the transitivity of verbs, that a parser might use to analyze a sentence or a generatror might use to construct one. Associated with the lexicon are programs that allow users to edit it and to construct and merge patch files to retain new lexical information across user sessions. Further information on these facilities appears in section I.6.8 of Appendix I.
While this lexicon has served us well, it is quite difficult to extend. The list-structure representation of subcategories has both limited our ability to add new subcategorical fields and hindered us from incorporating procedural information that would disambiguate word senses. The GEL Shell is designed to overcome these limitation. Uses for the shell are envisioned for applications beyond Relatus, however, and for that reason the GEL Shell is entirely independent.

Any computational lexicon intended to support natural language processing applications must provide facilities for the definition, acquisition, storage, and retrieval of subcategorial information. Unfortunately, one cannot simply state the appropriate subcategories, build the data structures for storage and retrieval of that information, define the appropriate procedures, and be done. A generic and extensible lexicon shell is needed, for the following reasons:

1. **Subcategories vary across languages.** Implementors require generic, language-independent tools for lexicon development.

2. **No consensus on subcategories has yet emerged.** Decisions regarding the necessity and appropriateness of subcategories depend upon the theoretical commitments of the implementor [Ingria, 1986; 1986a]. Implementors require lexical tools that are generic in the sense that they do not presuppose any particular linguistic theory.

3. **Implementors change their minds.** Implementors often discover that their original presuppositions of the appropriate subcategories were erroneous or underspecified, forcing them to respecify the contents of entire lexical entries. They need an extensible tool — one that will allow them to specify new subcategorical fields without requiring respecifications of the contents of previously specified fields.

The GEL Shell — a Generic, Extensible, Lexicon Shell — is designed to meet these needs. The GEL Shell is not a lexicon per se,
but a shell into which arbitrary lexical information may be stored, from which lexical information may be retrieved, and within which lexical information may be inspected and edited.

The GEL Shell is a meta-lexicon. It constitutes a collection of tools for storing and manipulating lexical information. It provides a set of functional abstractions and generic protocols designed to satisfy the diverse requirements that parsers, generators, and knowledge representation systems might impose upon a lexicon.

The genericity and extensibility of the GEL Shell make this range of applications conceivable. Rather than force applications programmers into a rigid, pre-defined conceptualization of appropriate lexical fields and protocols, the GEL Shell provides a kernel and a set of tools for defining and extending databases that encode lexical knowledge. This approach yields lexical structures of great flexibility.

In order to see how the GEL Shell is both generic and extensible, it is necessary to examine its components in some detail.

7.2 Shell Description

A LEXICON MANAGER is the top-level GEL Shell object. It supervises any number of LANGUAGE objects. Each language is composed of two classes of hierarchy — many lexical hierarchies and one non-lexical hierarchy. The non-lexical hierarchy encompasses those data structures that manage the information stored in the lexical hierarchies. Lexical hierarchies encompass data structures that store information about particular lexical items, i.e., phrases, words, categories, subcategories, senses, usages, etc.
7.2.1 The Lexical Hierarchy

The root of the lexical hierarchy is the LANGUAGE's LEXICON table, which associates tokens of a language with abstract (non-instantiable) objects representing its words, phrases, or morphemes. For convenience, we'll henceforth refer to these objects as WORDS. Below WORDS in the lexical hierarchy are one or more CATEGORY objects. Below each CATEGORY are zero or more SENSE objects. Below SENSES are zero or more USAGE objects. The lexical hierarchy is displayed in Figure 7-1.

```
Language
 |  
Lexicon Table
 |  
Words
 |  
Categories
 |  
Senses
 |  
Usages
```

NOTE: Plural nodes in the hierarchy indicate that many such objects may exist under its parent. Singular nodes indicate that only one such object may exist there.

**Figure 7-1:** The Gel Shell Lexical Hierarchy

Lexical objects are defined when the applications programmer invokes a DEFINE-LANGUAGE operation. This causes instantiable objects to be defined that inherit one of the abstract lexical objects. These newly-defined objects are specific to a language.
For example, when a language named SPANISH is defined using DEFINE-LANGUAGE, the instantiable object types SPANISH-WORD, SPANISH-CATEGORY, SPANISH-SENSE, and SPANISH-USAGE are defined. These inherit the slots and operations of the abstract object types WORD, CATEGORY, SENSE, and USAGE respectively.

These last four abstract objects are provided for applications programmers to specify operations across languages, while the language-particular object classes are provided for the specification of operations that pertain to particular languages. Internal GEL Shell operations are defined on the BASIC-WORD, BASIC-CATEGORY, BASIC-SENSE, and BASIC-USAGE objects, inherited respectively by WORD, CATEGORY-MIXIN, SENSE, and USAGE.

Applications programmers can define operations across categories within a language on any object class inheriting the CATEGORY object (e.g., SPANISH-CATEGORY). A DEFINE-CATEGORY form defines an instantiable object type for a particular category in a particular language. For example, when defining the PREPOSITION category in the SPANISH language, the DEFINE-CATEGORY form defines the SPANISH-PREPOSITION, which inherits SPANISH-CATEGORY. The applications programmer can then specify operations on SPANISH-PREPOSITIONs that pertain only to prepositions in Spanish. The DEFINE-CATEGORY form also allows applications programmers to specify object classes to "mix in" to these definitions their own abstract object types. Operations that pertain, for example, to prepositions across all languages might be specified on that "mixin".
WORD objects represent a single token. On each WORD object, a CATEGORY-TABLE associates category names (e.g., :ADJECTIVE) with CATEGORY objects. Each CATEGORY object has a list of SENSES that point to SENSE objects representing various senses of a word in a particular category. For example, the English word "pit" has at last two noun senses. It could be a hole in the ground or the core of a fruit. "Pit"'s verb sense would be stored under the verb CATEGORY object in the CATEGORY-TABLE of the ENGLISH-WORD named "pit".

Each CATEGORY object includes two slots that hold subcategory information. Most subcategory information can be represented as one choice among a limited number of possibilities. These are represented as one or more elements in a binary array. Other subcategorial information (e.g., synonyms or the plural of a noun) that cannot be compacted on a binary array are represented on the other array, which can represent any Lisp Datum.

Whenever an applications programmer defines a subcategory, a CATEGORY-EXPERT in the non-lexical hierarchy (described in section 7.2.2.2 below) allocates slots on one of these two arrays for the category. It then composes and compiles operations for setting and fetching subcategory information. Using the :PREDICATE-SPECS keyword to DEFINE-SUBCATEGORY (see Figure 7-3), applications programmers may specify names of predicates to return non-null values whenever a particular subcategory choice had been made for a particular lexical object. All these operations are defined on the appropriate instantiable WORD, CATEGORY, and SENSE objects for a particular language.
Since subcategory information must be specified in terms of a particular category, the information is stored on a CATEGORY object. Thus, the operations composed for WORD flavors require an additional argument — the category on which the subcategory is defined.

The automatically composed operations defined on SENSE objects differ in another way. It is possible for a SENSE of a word to override a subcategory value stored on one of the CATEGORY's arrays. Each SENSE object has an OVERRIDE slot that registers the name of a subcategory and its value for this sense of the word, when appropriate.

USAGE objects are not involved in the automatic composition of subcategorial set, fetch, and predicate functions. USAGEs encode examples of particular usages of particular senses of particular words, and might be used as a template for sense disambiguation. SENSE and USAGE objects include REFERENCE slots to retain an implementation-specific pointer to an entry-point in a knowledge representation of a description of the sense (the "meaning" of a particular word-sense) or an example of a sense (or a usage of the word in a particular sense).

For natural language understanding programs, the lexical hierarchy provides a convenient paradigm for the incremental resolution of ambiguity. Once the categorial ambiguities of a sentence are resolved (see section 3.5 of Chapter 3), the CATEGORY object can replace the WORD object in any representation. Once semantic or sense ambiguities are resolved, the SENSE object can replace the CATEGORY object in the representation. All relevant operations are defined on all three of these levels in the lexical hierarchy.
7.2.2 The Non-Lexical Hierarchy

From LANGUAGES, the non-lexical hierarchy forks off in two directions, as indicated in Figure 7.2. Down one branch is a patch manager. Down the other are category managers.

---

![Diagram of the GEL Shell Non-Lexical Hierarchy]

---

**Figure 7-2: The GEL Shell Non-Lexical Hierarchy**

---

The patch management facility allows end-users to retain across user sessions information they have provided to a lexicon. A patch manager supervises operations within and between various patch file objects. One patch file object supervises newly defined patches while the others supervise patch files that the user has loaded from a file.

Operations are provided that merge, intersect, and find the disjoint of two or more patch files. Other operations save the current patches to disk in a textual file and load patches into a lexicon during subsequent user sessions.
PATCH objects are organized in (alphabetized) balanced binary trees. Each tree is supervised by one of the patch file objects. Each PATCH object in the binary tree of current patches represents a word, phrase, or morpheme. Each associates a category with a CATEGORY–PATCH object, within which subcategory information is specified. In the lexical hierarchy, subcategorial information is represented in a compacted form to conserve space, as discussed above. In patch files, however, subcategorial information is stored in an uncompacted form. This allows human readers of a patch file to understand the contents of a patch readily, in order to detect any misspecifications.

Whenever a patch is made, the time of each entry and the name of the user specifying the entry are registered. This dependency information can be useful whenever a user's inaccurate understanding of the meaning of a subcategory causes incorrect information to be stored in patch files. Optionally, this information can be loaded in order to inspect or remove some or all information provided by a particular user, or before or after a particular time, etc. When merging patch files, lexicon maintainers can specify whose subcategory specifications to prefer whenever conflicts arise between patches.

7.2.2.2 Category Management

The LANGUAGE object associates categories (parts of speech), with CATEGORY–EXPERTS defined by a DEFINE–CATEGORY form. CATEGORY EXPERTS supervise the acquisition of subcategorial data from users and the composition of operations that store and access those data in the CATEGORY objects of the lexical hierarchy.
Each CATEGORY-EXPERT stands at the root of a hierarchy of objects (as indicated in Figure 7-2) that structure the information necessary to acquire subcategory information from end-users. If using a bit-mapped console, the CATEGORY-EXPERT and objects below it in the non-lexical hierarchy compose and invoke menus with which the user can specify the subcategory information for a particular token. The CATEGORY-EXPERT then registers this information in the appropriate objects of the lexical hierarchy and in the current patches.

If the user is logged in remotely, however, different operations are invoked. Subcategory acquisition questions are composed from the data structures in the non-lexical hierarchy, but the questions are asked linearly. Since the remote stream may not support menus and mouses, questions are instead formatted sequentially to the remote stream and the user responds with keystrokes.

Using a DEFINE-SUBCATEGORY form applications programmers can, among other things, (a) specify the order of subcategory acquisition queries, (b) group similar queries together on a menu (c) specify that different sets of queries are to use separate menus, and (d) specify constraints on the default value of a subcategory given the values of other subcategories or values returned by an arbitrary Lisp function. An example use of the DEFINE-SUBCATEGORY form is presented in Figure 7-3.
(define-subcategory
  :gender
  :language :english
  :category :noun
  :menu :properties
  :menu-group :basic
  :constraints ((:before :mass))
  :acquisition-string "What is the gender of "A? "
  :acquisition-args (ensure-singular self)
  :protocol (:choice masc fem either neuter)
  :predicate-specs
    ((masculine-p masc "Non-null if self is masculine.")
     (feminine-p fem "Non-null if self is feminine.")
     (masculine-or-feminine-p
      either "Non-null if self can be either gender.")
     (neuter-p neuter "Non-null if self is neuter.")
    )
  :help-function format
  :help-args
  (*standard-output*
   "Asks for the gender of "A. To answer this question, ~
   consider how one might refer to "A with a pronoun: ~
   ~If HE is appropriate, then answer MASC. ~
   ~If SHE is appropriate, then answer FEM. ~
   ~If either HE or SHE is appropriate, depending upon ~
   the context, then answer EITHER. ~
   ~If IT is appropriate, then answer NEUTER."
   (ensure-singular self)
   (ensure-singular self)))

Figure 7-3: The DEFINE-SUBCATEGORY Form

The :PROTOCOL keyword indicates that the user is to choose between
the four options listed there. Thus, when specifying the subcategories
for a particular noun, this subcategory will consume two bits in the
binary array associated with that noun. Other arguments to the
:PROTOCOL keyword allow specifications of subcategories that cannot be
expressed as choices (e.g., synonyms, verb forms, and verb
conjugations).
The :CONSTRAINTS keyword allows applications programmers to specify the order in which acquisition questions are asked. Similar :CONSTRAINTS keywords are available in the DEFINE-CATEGORY-MENU and DEFINE-CATEGORY-GROUP forms, which respectively allow applications programmers to specify the order in which menus appear and the order in which groups of acquisition queries appear within a menu.

Once the subcategorial information has been specified, the user selects among four options for disposition of the information.

1. "Done" signifies that all the choices have been made, and causes the information to be stored in the lexicon and the current lexicon patches.

2. "Help" causes a menu of query strings to appear asking the user to select one for which help is needed. The :HELP-FUNCTION for that subcategory (see Figure 7-3) is then applied to its :HELP-ARGS, which are evaluated within the environment of the object in the lexical hierarchy whose subcategories are being specified.

3. "Abort" causes the lexical object to be returned to its state before the acquisition routines began to run. The keyword :ABORT is returned, so that any program utilizing the lexicon can handle the situation gracefully.

4. "Wrong Category" signifies that the program invoking the subcategory acquisition procedures has incorrectly determined the category of a particular word. The lexical object is returned to its previous state and the keyword :WRONG-CATEGORY is returned.

A DEFINE-SUBCATEGORY-AUXILIARY-FUNCTION form allows applications programmers to specify an arbitrary piece of code to run after the content of a particular subcategory field is specified. Keyword arguments to DEFINE-SUBCATEGORY (not shown in Figure 7-3) specify the name and arguments of the auxiliary function. The arguments are
evaluated within the lexical environment of the object for which subcategories are being specified.

This feature is useful in many situations. Suppose, for example, that the user is queried about the number of a particular noun and responds that it is singular. If defined, the auxiliary function for the NUMBER subcategory would then run. It could then query the user for the noun's plural, perhaps offering a guess. Conversely, if the user answered that the noun is plural, the auxiliary function could query for the noun's singular.

7.3 Genericity

The GEL Shell is designed to be generic in seven ways.

1. **The GEL Shell is language-independent.** Different instantiations of a LANGUAGE object can store lexical information for a different language. The language objects are managed by the LEXICON MANAGER.

2. **The GEL Shell is theory-independent.** Top-level Lisp forms are provided with which applications programmers can define the categories and subcategories they wish to use. Applications programmers are free to mold the lexical data structures in a manner consistent with the assumptions and insights of any linguistic theory.

3. **The GEL Shell creates protocols for accessing lexical data.** The WORD, CATEGORY, and SENSE objects of the lexical hierarchy accept a common set of operations composed automatically be top-level forms invoked by applications programmers.

4. **The GEL Shell is not limited to syntactic subcategorization.** Since applications programmers define all categories and subcategories, any semantic or pragmatic fields may be added, consonant with the demands of particular applications.

5. **The GEL Shell is not limited to entries for single tokens.** Cross-token relationships are representable in the lexicon, along with a specification of the nature of their relationship. This phrasal feature allows lexical
recognition of idiomatic expressions and non-standard noun-phrase attachments [Becker, 1979].

6. **The GEL Shell is modular.** The objects defined by the GEL Shell are organized carefully to separate internal GEL Shell operations from the operations of applications programs, enhancing the understandability of end products. Additionally, operations may be easily defined for lexical objects of a particular language or across all languages.

7. **The GEL Shell implements a protocol for the acquisition of lexical information.** This protocol can alternate the procedures invoked to obtain lexical information depending upon the characteristics of the acquisition stream.

The GEL Shell allows applications programmers to extend lexicons gracefully. New categories may be added at any time, using the DEFINE-CATEGORY form. Each new category instantiates a CATEGORY-EXPERT that supervises subcategory acquisition for each token of that category.

During the development of natural language systems, implementors occasionally discover that additional subcategory information must be added to lexical entries. The GEL Shell allows these fields to be defined using high-level defining forms. System components (e.g., parser, generator, knowledge representation system) can access any new subcategory through the same instantiated objects, using the operations that the high-level defining forms define. This eliminates the need for major surgery on the lexicon or on any system that uses the lexicon whenever new subcategories are added.

Whenever new subcategory fields are defined for a particular CATEGORY object, the category's protocol of predicate functions, retrieval procedures, and object messages is automatically extended. When tokens for which subcategories had earlier been specified are
subsequently used, only the newly specified subcategory queries are performed.

When subcategories are deleted using an UNDEFINE-SUBCATEGORY form, the CATEGORY-EXPERTs mark them as deleted and they never again appear on acquisition menus. Prior to any binary or textual dump of the lexicon, the objects in the non-lexical hierarchy that control access to the arrays on the CATEGORY objects update themselves, the CATEGORY arrays, and the the relevant subcategory operations, eliminating those slots permanently.

7.4 Future Directions

At this writing, the GEL Shell has been completed and work has begun on constructing a prototype English lexicon. Future work envisioned for the shell are the implementation of a lexicon browsing and editing facility and a protocol for a network lexicon service, which would allow one machine at a site to serve lexical data to lexicons operating on other machines at that site.
8.1 The Problem

Temporal relationships are critically important in modeling narrative texts. Suppose, for example, that one was to model a narrative history of the Reagan Administration, and that, in 1990 Relatus is asked to represent the following sentences:

1. From January 20, 1981 to June 10, 1983, Alexander Haig was the Secretary of State.

2. From June 11, 1983 to January 20, 1989, George Schultz was the Secretary of State.

3. When the President was shot in 1981, the Secretary of State declared that he was in charge.

Now suppose that Relatus is queried, "Who declared that he was in charge when the President was shot?" Without some means of representing the temporal context in which Alexander Haig and George Schultz served as Secretary of State, Relatus would be hopelessly confused. At best, it could only answer that both Haig and Schultz declared themselves in charge. But this is clearly wrong. Haig made the declaration, and, since that information is present in the text, Relatus should be able to answer the query correctly.

Relatus requires a temporal component for other reasons as well. For example, temporal representation indispensable for any planning component. It would allow such a component to order actions and it would make possible the automatic collection of activities that co-occur into more general concepts represented complex processes
[Vere, 1983; Allen and Koömen, 1983; Allen, 1984]. Additionally, a temporal component can significantly enhance the efficiency of a planner [Vere, 1985].

To summarize the approach we plan for representing temporal relationships, the popular relations in sentences 1 and 2 would each be predicated with a dummy marker (e.g., AT-TIME) relating each instance of BE to a symbol denoting the appropriate time interval. These time-interval symbols are concatenations of numbers denoting the start-time and end-time of the particular interval at a particular level of resolution.59 By fiat, we define any interval which fails to span one unit at the chosen level of temporal resolution to be a "moment". An event beginning at the moment 462778 and ending at the instant 632563 would be denoted as the interval 462778–632563. Instantaneous events (those whose durations fail to span one unit at the level of resolution) would be represented as zero-length intervals.

Lisp code has been written for comparing two such intervals, returning the relation between any two intervals. For example, when handed as arguments the intervals 462778–632563 with 400000–700000, the algorithm returns :INFERIOR, meaning that the first argument is completely subsumed in time by the second. If the arguments are reversed, the value :SUPERIOR is returned. The full range of possible temporal relations is discussed in Section 8.2 below.

59 For a first cut, we will use the second as our level of resolution. This level is already present in time procedures provided in the standard Lisp Machine software. The method generalizes to any particular level of resolution, since any level can be represented by numerical values.
Using this trivial Lisp form, it would be possible to compare the time interval in the query (3) with those in (1) and (2). (1) is :SUPERIOR to the time interval in (2), so the subject of the relation predicated by that time interval, Alexander Haig, is the correct answer.

It is a rather simple matter to incorporate such temporal comparisons into the referential constraint-interpreter devised by Mallery [1987] for Relatus. But this by no means exhausts the work that must be done. Minimally required for the temporal component are an analyzer for temporal markings in syntax, a constraint-oriented temporal logic, and a temporal indexation mechanism.

Each of these modules is discussed immediately below. After providing rough characterizations of these modules, the paper goes on to present, in much greater detail, the characteristics of the temporal indexation component. Nothing in this paper is cast in stone. We remain open to the possibility that better models of each component may be discovered.

8.1.1 Syntactic Temporal Analysis

The syntactic markings that appear in texts must be recognized at parse time. The approach planned for this analysis is one that builds on theoretical work begun by Reichenbach [1947, 287–298] and extended by Hornstein [1977, 1981] and Yip [1986] on the English tense system, and on Woisetschlaeger's [1976] theory of aspect, which presents aspectual oppositions as knowledge about the sub-intervals of action.

The syntactic temporal analyzer would be implemented as a syntactic time expert associated with surface clausal agents in the
Relatus parser, discussed earlier. In addition to the relative clausal and inter-clausal temporal relations determined by implementation in the Relatus parser of the theories cited immediately above, the temporal analyzer will incorporate more precise temporal markings by analyzing temporal adverbials (adverb phrases and prepositional phrases) appearing in clauses. Together, these temporal markings will be included in sentential reference specifications as constraints on the referents of each clausal agent's verb phrase, handed to the constraint-interpreter [Mallery, 1987] at reference time.

Few time references are absolute, like "at three o'clock Tuesday afternoon" or "12 minutes from now". Most are fuzzy, like "yesterday morning" or "in about an hour". Fuzziness in time intervals will be represented by letting the start-times and end-times of time intervals be represented as intervals themselves. Start-times and end-times for non-fuzzy intervals are also represented as intervals, but zero-length ones. Symbols representing time intervals will be conservative, using the earliest possible start-times and the latest possible end-times. For their symbolic representation in Relatus belief systems, fuzzy intervals will be predicated by relations pointing to the intervals representing the range of their possible start-times and the range of their possible end-times.

The translation by the syntactic temporal analyzer of potentially complex temporal adverbial expressions into numeric representations including fuzziness factors requires a considerable amount of coding. This coding has already been accomplished for a wide range of typical temporal expressions, but much more remains to be implemented.
8.1.2 A Temporal Logic

Certain deductive inferences can be made with respect to temporal relationships. For example, if event A occurs before event B and event B occurs before event C, then event A occurs before event C. Temporal logics have been devised by Rescher and Urquhart [1971] and by McDermott [1982], among others. The logic should coordinate ordinal temporal relations with the interval-level representations of time intervals.

For example, a Relatus belief system might know that event A precedes event B, but it might know neither the precise end-time of event A nor the precise start-time of event B precisely. The syntactic temporal analyzer prefers the most conservative interpretation of a time interval. Ordinal relations between these time intervals are represented in Relatus belief systems by the verbal and prepositional relations which relate events (e.g., "precede", "succeed", "before", "after"). Whenever new information becomes available about either event, the other will be updated to reflect the new information.

The temporal logic component is envisioned provisionally as a set of activations (akin conceptually to pattern-invoked database daemons) that (a) deduce, when reasoning in Relatus belief systems, conclusions from temporal premises on the basis of a set of temporal axioms, and (b) progressively shorten the lengths of time intervals whenever new information becomes available about time intervals that are related ordinally to them.
8.1.3 Temporal Indexation

Some references cannot be satisfied tractably by simple comparisons between time intervals. Suppose that Relatus needed (in response to a query, for example) a list of all the time events that occurred during June of 1981. It would be prohibitively expensive to examine all time intervals in order to find only those that span, completely or partially, the time interval that denotes that month. One might also wish to construct worlds that are temporal subsets of a represented world. For this, some mechanism would be required for quickly finding all time intervals within a particular range. Additionally, when storing very large belief systems on disk, it may prove wise to compile the knowledge, representing it as sequences of state-changes over a time-line. When read back in, the original belief system would be re-created.

The implemented temporal indexation mechanism is designed to facilitate such operations in reasonable time. We will see below that we can do so in time proportional to the number of time intervals to be found plus the logarithm (base 2) of the number of represented time intervals.

The remainder of this document concerns the design and implementation of the temporal indexation mechanism. A prototype of the mechanism, implemented in Lisp, appears in Appendix II. To understand the approach adopted to temporal indexation, it will first be necessary to review the most successful of previous approaches to temporal indexation, indicating the shortcomings perceived in that model. Once this is done, the conceptual strategy developed to
overcome those shortcomings will be presented. Next, the first successful effort to implement an indexation mechanism based on that conceptual strategy will be described. While successful, the computational properties of search in this indexation mechanism were found to be undesirable. Finally, the alternative mechanism implemented in order to improve the performance of search will be presented.

8.2 Characteristics of Time Intervals

All of the 13 possible relationships between time intervals (including inverses) are presented in Table 8-1. These relationships are equivalent to those specified by Allen [1984], although some of the names differ from those used by Allen.

<table>
<thead>
<tr>
<th>NAME</th>
<th>INVERSE</th>
<th>RELATIONSHIP</th>
</tr>
</thead>
<tbody>
<tr>
<td>Superior</td>
<td>Inferior</td>
<td>XXXXXXXX</td>
</tr>
<tr>
<td></td>
<td></td>
<td>YYYYYY</td>
</tr>
<tr>
<td>Same Start Superior</td>
<td>Same Start Inferior</td>
<td>XXXXXXXX</td>
</tr>
<tr>
<td></td>
<td></td>
<td>YYYYYYY</td>
</tr>
<tr>
<td>Same End Superior</td>
<td>Same End Inferior</td>
<td>XXXXXXXX</td>
</tr>
<tr>
<td></td>
<td></td>
<td>YYYYYYY</td>
</tr>
<tr>
<td>Overlap</td>
<td>Underlap</td>
<td>XXXXX</td>
</tr>
<tr>
<td></td>
<td></td>
<td>YYYYYYY</td>
</tr>
<tr>
<td>Equality</td>
<td>Equality</td>
<td>XXXXXXXX</td>
</tr>
<tr>
<td></td>
<td></td>
<td>YYYYYYYYY</td>
</tr>
<tr>
<td>Starts at End</td>
<td>Ends at Start</td>
<td>XXXX</td>
</tr>
<tr>
<td></td>
<td></td>
<td>YYYYY</td>
</tr>
<tr>
<td>Disjoint Before</td>
<td>Disjoint After</td>
<td>XXXX</td>
</tr>
<tr>
<td></td>
<td></td>
<td>YYYYY</td>
</tr>
</tbody>
</table>

Table 8-1: 13 Possible Temporal Relations
In Table 8-1 the name in the first column denotes the relationship between interval X and interval Y, while the inverse in the second column denotes the relationship between interval Y and interval X. The simple interval comparison procedure, discussed briefly above, takes two time intervals as arguments and returns a value that indicates which one of these relations holds between them.

Based upon the 13 possible relationships between time intervals, one is tempted to build a data structure for each interval that would maintain pointers to all its nearest superior, inferior, overlap, underlap, etc. However, this approach is unworkable. The presence of overlaps and underlaps forecloses any hierarchical representation of time intervals, creating insurmountable difficulties.

---

Figure 8-1: Intervals Incapable of Hierarchical Expression

---

To illustrate, consider Figure 8-1, which presents four intervals graphically. Interval E underlaps interval D, which underlaps interval
C, an overlap of E. Interval B is superior to D and E and an underlap of C. Interval A is superior to B, C, D, and E. If all these relationships are explicitly represented by pointers between the intervals, there would be \([N(N-1)]/2\)=10 pointers, where N is the number of intervals so represented. As N increases, the time consumed in establishing these pointers quickly becomes prohibitive. The propagation of pointers must somehow be constrained.

One might want to constrain the propagation of pointers by pointing only to and from neighboring intervals. But how is this done? How shall the "neighbor" property be defined? In Figure 8-1 should interval E point to D or to C as its overlap? Should D point to B or C as its superior? Clearly, any decision would be arbitrary. For any procedure designed to intern intervals into such a network, the decision is likely to depend upon the order in which intervals are interned. Since circularities in this network would not be avoidable, procedures designed to collect all intervals within a given range could easily become confused and would not know when to terminate.

Allen [1981] constrains the propagation paths in his network by relating newly-interned intervals to reference intervals. Kahn and Gorry [1977] introduced reference intervals as a means of organizing a temporal database. For them, a reference interval would be some significant event, like one's birth, high-school graduation, or a trip to Europe. Other times would be interned along a time-line relative to these events of special significance. Allen's reference intervals are more discourse dependent.
Each time a relation is added, it is added with respect to a reference interval. The user of the system may set up the reference intervals as desired. In general, they probably will often reflect part of the during [superior-inferior] hierarchy, but may also reflect the semantic clustering of events [Allen, 1981: 223].

These reference intervals are themselves linked to their own reference intervals, so that a global network of temporal relations is maintained.

As noted above, however, there is no hierarchy, strictly speaking, of time intervals along the superior-inferior, or during, dimension. For example, consider again the intervals in Figure 8-1. Interval A is temporally superior to intervals B, C, D, and E. However, B is neither inferior nor superior to C. Interval E cannot be expressed as related to C or D in a hierarchy of temporal superiority/inferiority.

In general, for any pair of intervals related along the superiority/inferiority dimension, there are an infinite number of possible intervals that are inferior to the superior interval but neither inferior to nor superior to the inferior interval. So, unless one is willing to lose some intervals completely, it makes no sense to speak of a temporal during hierarchy.

Allen's reference intervals only diminish the propagation problem. They do not eliminate it. Reference intervals effectively partition a network of time intervals, limiting the propagation of pointers within a particular partition. However, overlap and underlap relations still prevent the construction of a network in which time intervals can be interned consistently and in which the intervals can
be located by procedures that do not cycle through portions of the network more than once.

In short, a network of time intervals related by direct pointers representing the temporal relations in Table 8-1 appears to be the wrong approach. To build a temporal network free of these problems, we need a complete reconceptualization of time.

8.3 Reconceptualizing the Temporal Network

Time can be characterized as a Cartesian coordinate system, the axes of which are the start-times and end-times of intervals. The system described below is presented graphically in Figure 8-2.\textsuperscript{60}

\begin{center}
\begin{tikzpicture}
\draw[->] (0,0) -- (4,4); \draw[->] (0,0) -- (0,4); \draw[dashed] (0,0) -- (0,0.5); \draw[dashed] (0,0) -- (0.5,0); \node at (0.5,0.5) {temporal octant}; \node at (4,4) {end}; \node at (0,4) {duration}; \node at (0,0) {start}; \node at (0,0.25) {midpoint};
\end{tikzpicture}
\end{center}

Figure 8-2: Cartesian Coordinate System for Time

\textsuperscript{60}Rit [1986] has independently reconceptualized time in a very similar manner.
Rotating the axes by 45 degrees, we find two axes that represent the durations and midpoints of intervals. Let the Y axis represent the end-times of intervals. Let the X axis represent their start-times. If we let the origin represent the hypothetical beginning of all time, then all time intervals would be represented as points in the temporal octant of Figure 8-2. No interval can appear below the origin on either the start or end axis, since it would have begun or ended before the beginning of time. This leaves only the upper right-hand quadrant. Since no interval can end before it begins, we are left with the half of this quadrant that is higher than the midpoint line in the end dimension. Notice that zero-length intervals, or moments, will always appear on the midpoint line.

For any one time interval, other intervals related to it by any of the 13 possible temporal relations in Table 8-1 will be located in an area of the space defined by the location of the interval and the midpoint axis. To illustrate, Figure 8-3 displays the interval T in a close-up view of the temporal octant. In Figure 89-3, the spatial locations of superiors, inferiors, overlaps, underlaps, and disjoints before and after are shown relative to the interval T.
Figure 8-3: An Interval's Relations in the Temporal Octant.

By definition, all superiors of any time interval must be located in that portion of the temporal octant higher in the end dimension and lower in the start dimension. Conversely, all of its inferiors must be located in the portion of the octant higher in the start dimension and lower in the end dimension. These areas are defined by lines parallel to the start and end axes that run through the point representing the interval. In Figure 8-3 these lines are BE and CF.
Points E and F are the points on those lines that also lie on the midpoint line. From E, we draw a line in the temporal octant parallel to the end axis. In Figure 8-3 this is line DE. From F, we draw a line in the temporal octant parallel to the start axis. In Figure 8-3 this is line AF. All of the interval's underlaps are found within the area bounded by AF, FT, BT, and the start axis (not shown in Figure 8-3). All of its overlaps are found in the area bounded by CT, ET, and DE. This area extends upward infinitely in the end dimension. All intervals that are disjoint before than the interval appear in the triangle bounded by AF, the midpoint line and the start axis. Intervals disjoint after the interval appear in the are bounded on two sides by DE and the midpoint line and extending infinitely upward in the end dimension.

Intervals that are equal to the interval of interest are represented as the same point in the space (T in Figure 8-3). The other possible temporal relations in Table 8-1 fall on line segments in Figure 8-3. These are listed in Table 8-2.
<table>
<thead>
<tr>
<th>RELATION</th>
<th>SEGMENT</th>
</tr>
</thead>
<tbody>
<tr>
<td>Same Start Superiors</td>
<td>CT</td>
</tr>
<tr>
<td>Same Start Inferiors</td>
<td>TF</td>
</tr>
<tr>
<td>Same End Superiors</td>
<td>BT</td>
</tr>
<tr>
<td>Same End Inferiors</td>
<td>TE</td>
</tr>
<tr>
<td>Starts at End</td>
<td>DE</td>
</tr>
<tr>
<td>Ends at Start</td>
<td>AF</td>
</tr>
</tbody>
</table>

Table 8-2: Temporal Relations Falling on Line Segments

By reconceptualizing the time network as points in one octant of a Cartesian coordinate system, where each point represents a time interval, all time intervals related to an interval in a particular way must be located in an area of the octant completely defined by the start-time and end-time of the interval, the start axis, and the midpoint line. This reconceptualization would not be helpful, however, if the amount of memory needed to store the coordinate system is excessive or if storage and access in the network consumes inordinate amounts of time (or processors, if implemented in a parallel machine).

8.4 Implementing the Temporal Network as a Coordinate System

Representing the temporal network as a two-dimensional array might easily consume far too much space. For example, if our level of temporal resolution were one second and we wanted our time network to
span one century, we would need an array greater in size than three billion squared. Each axis can be represented sparsely, however, by using binary search trees [Knuth, 1973: 422-447]. Two approaches were implemented, the second performing far better than the first. Both are described below, and the Lisp code that implements the second appears in Appendix II.

8.4.1 The First Approach

An n-dimensional binary-tree flavor mixin was written that could be inherited by agents representing temporal intervals (points in the temporal octant). Each temporal agent was made a node in four trees representing the start axis, the end axis, the midpoint line and the duration line. Essentially, the four trees represented projections from the points in the temporal octant to the two axes and the two 45-degree rotations of the axes.

Assuming that a binary tree is balanced (the total number of progeny for any node in the tree are roughly equal to the right and to the left), storage and access in the tree consumes time proportional to the logarithm of the number of nodes in the tree. It is highly unlikely that the temporal referents in an historical narrative, for example, would produce a balanced tree. Therefore, it was necessary to ensure that all four trees would be balanced. Two algorithms were implemented. One algorithm balanced the trees completely, so that, after balancing, a tree was in perfect balance. Unfortunately, the time performance of this algorithm is proportional to the size of the tree.
This algorithm was retained for use at full garbage-collection time, but the Aδel'son-Vel'skii and Landis (AVL) algorithm, described by Knuth [1973: 451-458], was implemented for use at intern time. The AVL algorithm runs in time proportional to the logarithm of the size of the tree, but it leaves the tree suboptimally balanced. That is, it allows the absolute value of the difference between the tree depths of one branch below a node to be greater than one. Only when the absolute value of the difference between the depths exceeds one will the balancing operation be performed. It is possible, therefore, that the maximum depth of the tree will exceed by one the logarithm of the number of nodes. This slight suboptimality in storage and access produced by the AVL algorithm is a small price to pay in exchange for the improvement it provides in the efficiency of balancing. Examples of a suboptimally balanced tree and its optimally balanced counterpart are presented in Figure 8-4.
Using the trees representing projections to the two axes and their 45-degree rotations, routines were written find all intervals related to a particular interval by one of the 13 possible relations in Table 8-1. Routines were also written, running in log(N) time, that determined which tree would find the intervals fastest. Because each tree represented projections of points on an axis or rotated axis, the procedures needed to filter out intervals that did not fall within the region of interest. For example, referring to Figure 8-3, the procedure finding all superiors of T using the tree representing the end axis had to filter out all of T's overlaps and disjoints after.

While this approach successfully found the correct intervals consistently, it did so in time proportional to the size of any one of
the trees (since all intervals are nodes in each tree, all trees are the same size). The operation that discovered which tree could perform the job fastest reduced the time needed by only a fractional amount. The time needed to find intervals within a region of the octant still rose monotonically with the size of the database, and performance would thus deteriorate too rapidly as the size of the database increased. A better approach was needed.

8.4.2 The Second Approach

A better approach would not rely upon projections to axes and their rotations, but would instead search areas of the two-dimensional octant directly. One way to perform such a search would be to represent the temporal octant as a quad-tree. Quad-trees are like binary trees, except that each node has four children instead of two. Each nodes represents a two dimensional space, while its children represent the four quadrants of that space. These quadrants are then further subdivided into quadrants, and so on, recursively. Assuming perfect balance, storage and access in balanced quad-trees is proportional to the logarithm (base 4) of the number of nodes. However, there is no known efficient algorithm for balancing quad-trees. Thus, the quad-tree approach was rejected.

Two dimensional spaces can also be represented using 2D-trees, which are well-suited for two-dimensional range searching [Sedgewick, 1983: 343-346]. 2D-trees are binary trees in which the dimension used

61Actually, since the time needed to find the next highest node in a binary tree with bidirectional pointers is proportional to the logarithm of the size of the tree, range searching in these trees is proportional to N log(N).
to determine at intern-time whether to go right or left from a particular node depends upon the depth of the tree. For interning intervals into a 2D-tree, the end-time of the interval might be used as the determinate at even-numbered depths while the start-time is used at odd-numbered depths. The resulting tree successively bifurcates the two-dimensional space, allowing procedures to find areas of the space in time proportional to logarithm of the size of the tree plus the number of nodes to be found [Sedgewick, 1983: 346].

To illustrate the use of 2D-trees, consider the following intervals, generated randomly and interned into a 2D-tree in the following order 1: 446–971, 2: 15–633, 3: 415–817, 4: 28–840, 5: 90–404 6: 42–382, 7: 181–388, 8: 316–942, 9: 381–448, 10: 589–995. The tree would appear as presented in Figure 8-5.

---

62 Sedgewick provides no analysis for or proof of this. However, experience with the 2D-trees indicates that such a proof would be possible. Range-searching procedures search only those pathways within which a node satisfying the range boundaries could possibly exist, ignoring other pathways. In the implementation described below, the practical time consumed in range-searching is further shortened in many cases (but not in the worst case) by exploitation of nodes' pointers to the highest and lowest values below themselves in the trees.
Figure 8-5: 2D-Tree for Randomly Selected Intervals

By alternating the dimension along which comparisons are made when interning into the tree, the space represented by the tree is successively bifurcated. At any depth of the tree, the space is divided into 2D subspaces, where D is the depth of the tree. For example, the tree in Figure 8-5 divides the temporal octant as indicated in Figure 8-6. Note that all intervals fall necessarily into the temporal octant.
Figure 8-6: Temporal Octant as Divided by Tree in Figure 9-5.

Unfortunately, the 2D-tree in Figure 8-5 is grossly out of balance. As a result, grid searching using this tree is not likely to
be efficient. The AVL algorithm will not work for 2D-trees. Applying it would violate the alternating-dimensions semantics of the tree. Therefore, an algorithm for balancing 2D-trees in \( \log(N) \) time was devised. To explain the balance algorithm, it is first necessary to explain the operations that occur when an interval (or node in the 2D-tree) is stored.

### 8.4.3 Storage Procedures

Like the AVL algorithm, this algorithm requires that each node maintain the depth of its progeny on both its left and right branches. In order to balance 2D-trees, however, each node (representing an interval) in the tree must maintain two additional pointers. These point to nodes in the tree that represent the highest and lowest intervals below in the dimension opposite to the current node's dimension for node comparison.

When an interval is recursively interned in the tree (beginning at the root), if an interval is already present in a position, a comparison is made to see whether the interval being interned should be stored among the left or right progeny of the already-interned interval. Nodes at odd-level depths use end-times to make this decision. If the end-time of the newly-interned interval occurs before (is less than, since times are represented as numbers) the end-time of the already-interned interval, the new interval is sent to the left-hand child for storage. Otherwise, it is sent to the right-hand child. Nodes at even-level depths do the same, except that they use the start-times of intervals as the basis for comparison. Whenever no
child exists in the appropriate direction, the newly-interned interval becomes that child.

During this intern procedure, each interval along the intern pathway checks the node being interned against the two nodes to which it points as the nodes with the highest and lowest values in the dimension opposite to the dimension it uses for comparison. If the existing node does not point to any node, the newly-interned node becomes the highest or lowest node below it. If the existing interval does point to a node, the node pointed to is compared to the node being interned. If the latter is higher than (or lower than) the node pointed to, the existing node is made to point to the newly-interned node instead.

After the recursive intern of an interval, information pertaining to the depth of each node's progeny is updated recursively along the storage pathway (whose length is approximately \( \log(N) \), since balance is maintained on every intern of a node), from the newly-interned interval on the fringe to the root of the tree. If, at any node, the absolute value of the difference between the depth of the left progeny and the depth of the right progeny exceeds 1, the node is out of balance and the balance algorithm is run. Since the last operation of the balance algorithm recursively updates the progeny depths, the updating of progeny depths is terminated before balancing. The later call to the procedure ensures that the unchecked nodes toward the root will ultimately be checked, and, if necessary, balanced.
8.4.4 Balance Algorithm

The balance algorithm replaces any imbalanced node with a node below it, and re-stores the imbalanced node inside its replacement. For example, if the maximum depth of the left progeny of a node with an odd depth in the tree exceeds by two the maximum depth of the node's right progeny, the replacement node will be the node with the highest end-time below the node and to its left. This node will either be the imbalanced node's left child, or the node to which the left child points as its lowest below, depending upon which has the latest end-time. Since the depth of the child will be even, its pointer to the lowest below will be the node with the lowest end-time below it. No other nodes in the sub-tree need be examined. Replacement nodes for other imbalance conditions are handled analogously, depending upon the depth of the imbalanced node and the direction of imbalance, as indicated in Table 8-3.

<table>
<thead>
<tr>
<th>Depth</th>
<th>Direction</th>
<th>Replacement</th>
</tr>
</thead>
<tbody>
<tr>
<td>odd</td>
<td>left</td>
<td>Highest end-time below left.</td>
</tr>
<tr>
<td>odd</td>
<td>right</td>
<td>Lowest end-time below right.</td>
</tr>
<tr>
<td>even</td>
<td>left</td>
<td>Highest start-time below left.</td>
</tr>
<tr>
<td>even</td>
<td>right</td>
<td>Lowest start-time below right.</td>
</tr>
</tbody>
</table>

Table 8-3: Replacements for Imbalanced Nodes

Once the replacement node is found, it is then spliced out of the tree and reinserted in the imbalanced node's position. The imbalanced
node is then reinserted in the tree by interning it underneath the replacement node. It will intern itself down the fork in the tree opposite to the fork from which the replacement was found.

Once the formerly imbalanced node is reinserted the depth-adjustment procedure is re-invoked. If all nodes are now in balance, this procedure will traverse the tree upward through the replacement node (which is why we were able to terminate depth adjustment when we found the imbalance) to the root. It is possible, however, that any one of these nodes, including the replacement node, may be out of balance. In that case, the depth-adjustment procedure is terminated and the balance procedure is reinvoked.

The perspicacious reader may have noticed one or two problems with the balance procedure. First, any node between an imbalanced node and its replacement may point to the replacement as the highest or lowest node in the opposite dimension. Second, the replacement node may not be located on the fringe of the tree, which means that an entire subtree of the node would require reformulation. This reformulation would consume time proportional to the number of nodes in the tree. Both of these difficulties are easily overcome.

8.4.4.1 Pointer Adjustment

Once the replacement node is spliced out of its position in the tree, a procedure walks up the tree from the replacement's parent to the imbalanced node, checking the pointers maintained by each. Because the replacement node will be moved to a position higher in the tree, any pointer to the replacement must be corrected to reflect the actual nodes that are lowest and highest in the opposite dimension. The entire
subtree below need not be searched to find the appropriate node for this slot. It can only be located in a finite number of places. It must be either (a) one of the children of the node in question, (b) one of two of its grandchildren (depending upon whether the highest or lowest value is sought), or (c) one of the nodes to which a grandchild points as the highest or lowest in the opposite dimension. These are compared, and the node that best satisfies the requirement (highest or lowest in a particular dimension), replaces the replacement as the lowest or highest below for that node.

In the worst case, there can be no more than six alternatives. If the imbalanced node is at the root, the replacement is at the fringe of the tree, and all nodes along the pathway point to the replacement (and must therefore be corrected), there can be no more than \( 6 \log(N) \) comparisons. Of course, in practice, the number of comparisons will usually not approach this number.

### 8.4.4.2 Replacements Not at the Fringe

The balance procedures described above assumes that the replacement node is at or near the fringe of the tree. Specifically, the balance procedures assume that the node is terminal, or that it has no progeny beyond one or two children. In these cases, reformulation of any tree structures below the replacement node is trivial. Occasionally, however, the replacement node will be found much higher in the tree (although always below the imbalanced node). It is unfeasible to splice the replacement out of such a location. One or the other of its two children would have to replace it in the tree. The remaining progeny below the replacement's replacement would then
decrease by one their depth in the tree. Since the relative positions of nodes in the tree are determined by the values of intervals in different dimensions, that whole subtree would require reformulation.

This difficulty is readily overcome by performing two balances instead of one. Whenever the replacement node is found at a level in the tree higher than the fringe (or one level above the fringe), the replacement node is balanced, even though it may not be out of balance. As a result of this secondary balance, the replacement node is lowered to the fringe of the tree, and the balance procedure can be recursively invoked on the node that was originally out of balance.

8.4.5 The Benefits of Balance

By balancing the tree upon the intern of each new node, the 2d-tree in Figure 8-5 assumes the shape displayed in Figure 9-7. Without balancing, it is entirely possible that interning any new node and performing any other operation (like range-searching) on the 2d-tree would perform in time proportional to the number of nodes in the tree. Balancing ensures these operations will be performed in \(\log(N)\) time.
Because the tree is balanced, the progressive divisions of the two-dimensional space, as defined by the tree, are much more equally divided. At each level of division in the temporal octant, each subdivision contains approximately the same number of nodes. Figure 8-8 subdivides the temporal octant defined by the tree in Figure 8-7. Compare this division of the octant with the unbalanced division in Figure 8-6. on page 257. As a result, range searching procedures operate more efficiently when the 2D-tree is balanced, homing in on the appropriate space more quickly.
Figure 8-8: Temporal Octant as Divided by Tree in Figure 9-7.

One unexpected result has emerged. The pointers maintained by each node in the tree to the nodes below that have the highest and
lowest values in the dimension opposite to that node's dimension of comparison serve as a useful constraint on range-searching.

For example, consider the tree in Figure 8-7 on page 264. Suppose that we are searching for intervals whose lowest end-time is 650 and whose highest end-time is 800. The range-searching procedure ignores the right fork of the root node, 415-817, because the end-times of all nodes down that fork must exceed 817 and are therefore outside the range of interest. Search then proceeds to the left child, 42-382. From this node the direction of search would ordinarily depend on the start-time range of interest. However, this node knows that the interval with the highest end-time below it in the tree is 15-633. Thus, no node can exist which satisfies the end-time range specification, and the procedure can terminate immediately, without searching any further.

8.5 Conclusions

Representations of temporal relationships in a two-dimensional space are far more natural than graph representations that relate temporal intervals by means of direct interval-to-interval pointers. It is possible to represent the space using binary trees as representations of dimensions in the space reflecting the start-times, end-times, and possibly also the durations and midpoints of intervals. However, range searching in this representation takes time proportional to the size of a tree times its logarithm (base 2).

All operations, including interning and range-searching, become far more tractable when time is represented in a 2D-tree, using the start times and end-times of intervals as dimensions. Since only one
tree is created, a considerable space savings is also gained. The tractability of this approach is enhanced further by balancing the 2D-trees. Unlike another alternative approach, quad-trees, it is possible to balance 2D-trees, in log N time, thereby enhancing the performance of all operations. Procedures for balancing a 2D-tree were developed and described in some detail.

The balance procedures require the maintenance by each node of the highest and lowest nodes below it in the dimension opposite to that node's dimension of comparison. This information also serves as a useful constraint on range-searching, thereby speeding this important operation for many cases. Balanced 2D-trees are therefore strongly recommended as the appropriate method for indexing temporal intervals.
9.1 Intention and Intension

The recent decline of the positivist hegemony over the academy presents the opportunity for the full exploitation of political texts and utterances as data for political analysis. Under positivistically oriented behaviorism, political analysts could not understand linguistic productions as disclosures of (or as camouflage for) constructs within the mind of an individual political actor. To be sure, political analyses of textual data were always possible, but the very possibility of consensual redemption of validity claims arising from any such analysis had, until recently, been foreclosed.

As Alker [1975: 188-189] observes, post-positivistic research into the beliefs and intentions of key individual political actors can provide us with insights into how they understand, and thus also shape, their political worlds. While this sort of "intentionalist assessment", as Alker calls it, certainly has merit, we need to advance further. We need an assessment that is intensionalist as well as intentionalist.

As discussed in Chapter 2, the Conceptual Dependency techniques that Abelson et al. deploy to construct their Goldwater models presume the universal validity of particular content categories. Conceptual

---

62See Alker [1975] for references to these studies. The production and analysis, by Robert P. Abelson and his associates, of Barry Goldwater's belief system represent the earliest efforts in this direction.
Dependency theory can thus be seen as a proposal for a universal description language.

As an effort to implement as computational environment for hermeneutic analyses, we cannot presume a rigid set of content categories in Relatus. The textual interpretations of any model must therefore be grounded in its previous interpretations, particularly in the necessary, sufficient, and _ceteris paribus_ conditions of conceptual constructs believed by the actor, not in the pre-programmed interpretations of an implementor. This places an additional, hefty burden on our efforts.

As discussed in section 3.5, even the relatively simple processing decision of prepositional phrase attachment requires reference to the underlying conceptualizations and categorizations of a text interpreter. The lexical databases proposed for encoding "subcategories" that inform those decisions constitute yet another category scheme that, cannot be presupposed in any computational hermeneutic. While the prototypical lexicon shell described in Chapter 7 circumvents this difficulty somewhat by providing for sense-specific subcategorial specifications, ultimately such decisions will require reference into the (simulated) beliefs of the interpreter -- into the encyclopedic store of a simulated life-history.

No simulated life-history, however can immediately be generated. In order to construct these models, analytic components that presuppose their existence, e.g., parsers, lexicons, generators, etc., must first exist. We must therefore bootstrap a system that meets the hermeneutical requirements of interpretive validity.
Lest anyone believe that the exegesis of these difficulties reveals the project to be quixotic, we fully expect significant results as we progress. Beyond the initial application of Relatus in detecting in analyzing the phases of intentionally sensitive prisoners' dilemma protocols [Hurwitz, et al., 1987], other politically relevant analytic modes currently approach feasibility. The capacity to model the relationships that sentences convey allow us to declare in English the relationships that one might observe in any social setting. These social relations, whether they be economic relations, personal ties, affect states, expressive of beliefs or desires, or some combination of these, then become amenable for systematic, replicable analysis.

9.2 Political Schemata

In the course of developing a system for text analysis using political texts as data, we should gain, as a positive externality of the effort, a more well-grounded notion of political schemata. Schema theory, proposed by Axelrod [1973, 1976] (following Abelson et al.), is an emerging body of literature in political and social psychology that couches explanations of ideology in the coherence of individuals' beliefs, attitudes, and opinions.63

While the appearance of this literature advances political psychology, it has quite a distance to go. As Jervis [1986] notes, the political cognition literature has thus far failed to integrate

63For recent introductions to and examples from this literature, see [Fiske and Taylor, 1984] and Lau and Sears, 1986].
relevant theories, such as Festinger's cognitive dissonance theory and Kelley's attribution theory.

A final problem is that social cognition is an outlook, not a theory. It consists of essentially untestable assumptions and a series of somewhat loosely connected demonstrations and propositions. But there is not enough in between these two levels. Thus, the approach remains disturbingly ad hoc and does not lend itself to the deductive [sic] generation of important hypotheses. In part, scholars have followed this path as a reaction to earlier lines of investigation, such as [Heider's] consistency theory, which gained a high degree of unity at the cost of descriptive accuracy. In addition, the lack of coherence may partly be a function of the relative youth of the approach; it may become more of a theory as more work is done [Jervis, 1976: 323].

Jervis' observation that political cognition at present is too ad hoc to allow hypothesis generation is right on the mark. But the ad hocness lies not merely in the fact that earlier cognitive approaches of interest to political psychologists have not been integrated. The notion of "schema" itself seems radically underspecified in this literature.

In Axelrod's [1976] study of the decision-making behavior of the British Eastern Committee, political schemata are represented as "cognitive maps", or complex concepts linked by causal relations indicating whether one concept has a positive (increasing) or negative (decreasing) causal effect upon another. For example, two of Axelrod's concepts are:

1. Allowing Persians to have a continued small subsidy.
2. Ability of Britain to put pressure on Persia.

These two concepts are related by a positive causal effect. That is, (1) increases (2).
While it is surely plausible that members of the British Eastern Committee believed that British subsidies to the Persians afforded them a lever with which to manipulate the behavior of the Persians, this level of analysis is quite high. Suppose, for example, that the concept of "pressure" differed in meaning among members of the Committee, or that a "small" subsidy for one member was for another member quite large. Decisions by members of the Committee might hinge not only on the causal relations of such concepts, but also on the meanings, for each member, of the terms used to describe the concepts. These are rather minor points, but they do indicate that, for some analytical purposes, we require more detailed models of political "schemata".

In the bulk of this literature political schemata are not modeled at all. Rather, they are generally postulated as constructs that purportedly explain correlative survey research findings. These explanations, of course, do not fall out of the sky. The analyst supplies them, but on the basis of categories that are already specified in the instruments used to gather the data for analysis. There is a curious circularity here. The analyst prespecifies the questions to be asked and, unless the answer is open-ended the legal response categories. The subjects dutifully provide the answers. The data are correlated, and any observed correlations are cast as evidence for the existence of this or that cognitive schema. The questions chosen by the analyst and the response options provided by the analyst already influence strongly the nature of these "schema". Missing from
the nascent political cognition literature is any grounding for the constructs employed in explanations -- for the schemata themselves.

In order to satisfy Jervis' call for a theory from which testable hypotheses can be "deduced" he means "abduced", of course), we require a more precise theoretical account of "schemata". The best way to go about this, it seems, would be to construct and test computational models that might plausibly, based on some defensible, prior theoretical commitments, represent mental "schemata". Proposals that are advanced for the nature and function of cognitive schemata can best be compared and tested when implemented computationally.

Any such proposal must of course meet some minimal performance criteria in order to be considered. It must be capable of encoding models of the relations we observe visually and it must be able to record the relations we observe when confronted with an utterance. It must perform the quasi-logical operations we humans perform when we draw inferences from our visual and linguistic perceptions, and these operations must be performed tractably, breaking down only where humans break down. From such proposals a vocabulary will emerge that will allow us to construct more comprehensive and more comprehensible explanations for the behavior we observe when humans process information -- political and otherwise.

9.3 A Reflective Social Scientific Practice

As reflected in its title, this thesis involves three interrelated topics -- language, politics, and method. It primarily concerns the development of methods for the analysis of political language, but the three topics are reflected in subthemes that emerge in various places
within the document. Some subthemes are treated more prominently than others, of course, but a consideration of the variety of ways in which these three topics combine seems a useful means for conclusion.

First consider the language of politics. This theme reflects the primary goal of the Relatus project -- the development of methods with which to model and analyze the deliberations of groups, political propaganda, political debate, political bargaining -- in short the whole variety of conversation types in which the political presents itself. By building and analyzing models of these discourses and of the cognitive structures they disclose, we hope to understand better the cognitive incommensurabilities that arise across interpreters in the opposing camps of political conflicts. There is no suggestion of course that conflicts can be understood without reference to their material bases. Rather, the suggestion is that community conflict is more readily resolved when the incommensurabilities that develop as a result can be brought to the fore, analyzed, and understood, and when conflicting groups can be informed of the history of conceptualizations that ground the seemingly incomprehensible interpretations of their political adversaries [Alker, Bennett, and Mefford, 1980].

Chapter 4 itself constitutes a discursus on the politics of linguistic method. The methods that distinguish formally inclined generativists from empirically inclined interactionalists seem remarkably similar to those that distinguish formally from empirically inclined political scientists. The formalists in both disciplines posit unrealistic simplifying assumptions from which very powerful explanations have emerged. In their critiques, the empirically
inclined typically point out the unreality of these simplifying assumptions and demand that formal results be viewed with suspicion until those assumptions are relaxed. Formalists, for their part, decry the weakness or, in some cases, the absence of theory informing the work of their empirically oriented critics. The contenders in both disciplines seem to have achieved uneasy stalemates. The formalists cannot relax their assumptions to the degree demanded by the empiricists without forfeiting explanatory power. Conversely, efforts by the empiricists to induce powerful theoretical constructs are continually foiled by the complex, enigmatic, subjective humans whose behavior, linguistic or political, they endeavor to explain.

This last point raises the theme of the linguistic politics of social-scientific method. Theory development in the social sciences is hindered by the fact that the social scientist, unlike the natural scientist, must engage in a conversations with the objects of inquiry. Social scientific inquiry is itself a political phenomenon in which the subject of science and the object of science must negotiate meanings and understandings. The social scientist cannot without profoundly distorting effects, unreflectively apply the techniques of control available without cost to the natural scientist.

Deduction, induction, and abduction establish relations between statements that are in principle monologic. It is possible to think in syllogisms, but not to conduct a dialogue in them. I can use syllogistic reasoning to yield arguments for a discussion, but I cannot argue syllogistically with an other. Insofar as the employment of symbols is constitutive for the behavioral system of instrumental action, the use of language involved is monologic. But the communication of investigators requires the use of language that is not confined to the limits of technical control over objectified natural processes. It arises from symbolic interaction between societal subjects
who reciprocally know and recognize each other as unmistakable individuals. This communicative action is a system of reference that cannot be reduced to the framework of instrumental action [Habermas, 1968: 137].

The normative side of Habermas' theory of communicative action is the theory of communicative competence. Here we touch on the politics of language. According to this view, in the very act of engaging in communication we anticipate the possibility of coming to a rational consensus — a consensus based not on irrational constraints on the discourse process, such as coercive physical force or internalized psychic constraints, but on the force of the better argument. If such a consensus were not possible, we would have little reason to enter into discourse. Habermas [1979] raises this observation to a point at which it elaborates and extends the principle of publicity in democratic theory. Our anticipation of rational consensus in discourse discloses to ourselves our emancipatory interest in an ideal speech situation — a community in which collective decisions are made in a discourse open to all and free from constraints on communication. Of course, the ideal speech situation remains only an ideal. But it need not be attainable in order to provide social and political direction.

In order to approach the ideal speech situation, it is necessary to attend to the communicative competence of possible discourse participants. The the degree to which the ideal is approached, then, necessarily corresponds to the degree to which all potential discourse participants can communicate free from internal and external constraints. Freedom of speech would thus be maximized while cognitive constraints on communication (based for instance in low self-esteem engendered by and enforced through domination, subtle or otherwise)
would be minimized. It means also that social policies would be directed toward ensuring that children enjoy the opportunity to achieve a level of communicative competence at which they can eventually become effective, adult participants in the discursive process of social will-formation.

This last desideratum would hardly be needed, however, on the theory that children acquire language simply by setting the parameters of principles that are innately provided. Chomsky, of course, speaks only of syntactic competence while Habermas speaks of communicative competence, the latter involving semantic and pragmatic competence as well. But these distinctions make little difference here. One need only examine whose children enter what social strata to recognize that great disparities exist with respect to these competencies. And one need only compare the syntactic abilities of persons from varying class backgrounds to recognize that syntactic competence is no exception.

Finally, we arrive at the subject of the language of method. The construction of computational models capable of incorporating all the relationships that can be expressed in language can provide great dividends. We have already noted that the ability to incorporate effable relationships into such models stands as evidence for their cognitive reality. As this work proceeds, we begin to develop a sense of how we process this information, how we form hypotheses, how our conceptual frameworks simultaneously disclose some aspects of reality and blind us to other aspects. We begin to develop a vocabulary for expressing not merely our thoughts, but also our thoughts about our thoughts. We disclose to ourselves the cognitive methods we employ to
disclose cognitions to ourselves and to foreclose them from ourselves. As a result, we become more reflective, as scientific practitioners and as citizens, and as people. And this by itself, apart from any other benefit, makes all the effort worthwhile.
References


Alker, Hayward R., Jr., "Bit Flows, Rewrites, Social Talk: Towards More Adequate Informational Ontologies," paper presented to the Todai Symposium, the centennial symposium of the University of Tokyo, October 1986. Revised version of May 6, 1987, MIT Department of Political Science.


Habermas, Jurgen, *Knowledge and Human interests*, (Boston; Beacon Press, 1971).


Kubie, L., "Blocks to Creativity", in [Mooney and Razik, 1967], pp. 33-42.


Mallery, John C., Roger Hurwitz, and Gavan Duffy, "Hermeneutics: From Textual Explication to Computer Understanding?", Massachusetts Institute of Technology, Artificial Intelligence Laboratory, Memo No. 871, 1986.


Winograd, Terry, and Fernando Flores, *Understanding Computers and


Appendix I
Relatus Overview

by

Gavan Duffy and John C. Mallery

Abstract

This paper introduces the Relatus natural language system, an implemented artificial intelligence system designed for manipulating cognitive models constructed from English texts. We first situate the Relatus research effort within the context of the social-scientific needs that motivate the research. We then explain the how these motivations translate into the general design heuristics that have guided the development of Relatus. At the most abstract level, we intend Relatus to be an experimental tool for representing and analyzing cognitive processes associated with political cognition. We see political cognition as grounded in the historical experiences and cultural traditions of political actors. Thus, an adequate modeling tool must be able to model the differing interpretations of a single text which vary according to differing background knowledge and differing inferential regimes. This hermeneutic consideration means that text alone is not data. Only with the addition of interpretation do we obtain the data — knowledge-based models — for study and experimentation. This represents the key contribution of the natural-language processing approach: it makes interpretive issues an explicit part of the modeling task rather than hidden presuppositions buried in an unanalyzed precoding. Thus, natural-language modeling provides technology for formal and reproducible analyses of politically relevant discourse and associated political cognitions. After reviewing some of these theoretical foundations, we overview the system's major components and discuss their functions.

I.2 Introduction

The Relatus system is an implemented artificial intelligence (AI)
system for representing and analyzing English texts. The capability to construct "word models" from texts prepares the development not only of analyses of the structure of political discourses, but also for simulating cognitive processing (such as learning and reasoning) based on the semantic contents of those texts. The system already has a knowledge acquisition capability. It can incrementally construct knowledge representations of text by parsing sentences and performing intrasentential and intersentential reference. Presently, the system is restricted to single-sense word usage in literal and explicit texts. Yet the acquired knowledge can be deployed to answer various classes of literal and explicit questions. These abilities comprise a solid foundation upon which to build sophisticated reasoning systems.

In its major processing cycle, the system maps written texts into a dynamic knowledge representation that captures the referential structure of a text. This representation lends itself to the regeneration of surface texts and to inferential operations associated with common-sense reasoning. Relatus can presently analyze literal and explicit texts and map the analysis into a felicitous representation. Current efforts are directed toward the development of common-sense learning and reasoning capabilities based on those representations.

I.2.1 Modeling Political Cognition

The Relatus project aims to build an environment for the

---

56 The system is implemented in ZetaLISP and runs on Symbolics 3600-class lisp machines.
systematic study of political cognition. Political cognition has two fundamental features:

1. It is grounded in "procedural rationality" [Simon, 1985], a non-stochastic, symbolically-mediated, functional account of human reasoning processes, where these reasoning processes are conceived as "physical symbol systems" [Newell, 1980].

2. It is fundamentally a linguistic phenomenon, centered on interpretations of the meanings and intentions of social actors [Gadamer, 1960; Alker, 1979a; Habermas, 1981; Dallmayr, 1984].

Our approach recognizes that Man is not merely *homo faber*, but is, perhaps more fundamentally, *homo loquens*. People live in linguistic communities and exchange information through language. Significant action can be treated much like the linguistic statements found in texts, because both convey meanings and a subject to interpretation that reveals motivating intentional structures. By building representations of intentionality in addition to belief, it becomes possible to model social action [Ricoeur, 1971; Searle, 1983]. While political cognition may be studied using traditional, non-computational, "pencil-and-paper" approaches, the level of complexity, systematicity, and reproducibility that can ultimately be realized by such intuitive methods is limited. Thus, the Relatus project is a basic research effort, designed to yield formal tools for modeling and studying the kinds of cognitive processes involved in political action.

1.2.2 Locating Relatus in Research Traditions

The social scientific motivation for the Relatus project derives from six earlier modeling traditions:


5. Control Theoretic Models Of Social Systems [e.g., Forester, 1971; Meadows, et al., 1972; Meadows, et al., 1974; Mesarovic & Prestel, 1974; Cortes, Przeworski & Sprague, 1974; Alker & Tickner, 1977; Meadows & Robinson, 1985].


While seemingly disparate, these traditions are united by a shared concern for inherently qualitative methods of representation and analysis of social phenomena [Alker, 1975, 1978, 1979a]. Each is targeted at blind spots left by the stochastic methods that have been heralded by some as the hallmark of "a scientific social science." Stochastic methods effectively perform a type of extrapolation between the available data points that can reveal outlines of complex systems and can predict system behavior based on inertia. But these successes also inherently lead to poor performance on the low-probability non-linearities that often account for major system transformations.

Information processing approaches, content analysis, feedback models, cognitive mapping, cognitive models, and precedent-based models
each attempt to overcome this weakness, but in different ways. Information processing approaches attempt to model social phenomena through the lens of problem-solving and planning behavior. Information processing models aim at elucidating goal-seeking behavior using notions of "bounded rationality" derived from structural models of general problem solving [Newell, Shaw, & Simon, 1957, 1959].

Content-analytic techniques have proven helpful in certain macrosociological applications, where general sociopolitical value orientations are the research focus [e.g., Namenwirth, 1973; Weber, 1977], and in practical political applications, where the research orientation is discovery of fifth columnists or an enemy's intentions [Lasswell, Leites, et al., 1949], not scientific discovery. While computerized, quantitative content-analytic techniques [e.g., Stone, et al., 1966] can provide information regarding the kinds of value orientations expressed in particular textual corpora, they cannot provide information about the direction of value orientations [Duffy, 1986a]. For political texts, this is of critical importance.

While content-analytic methods seek to uncover the affective side of word usage based on thematic content, cognitive mapping approaches attempt to capture affect based on "cognitive consistency" or affective associations. Cognitive modeling approaches reflect the realization that structural model of natural language processing and interpretation must move beyond superficial correlational belief attributions [Alker, 1975, 1979a].

Building from control theory, feedback modelers deploy differential equation models, thereby representing non-linearities in
complex systems, accounting for instability and stability. Although feedback modelers initially placed considered weight on the numerical values in their equations, they have since come to emphasize the qualitative relationships of positive and negative feedback, morphogenesis and morphostasis. Thus, while these methods involve numerical solutions to differential equations, their primary goal is the articulation of the structure and signs of feedback relationships. Thus, they address a predominantly qualitative problem involving discrete state changes.

For our purposes, the information processing models of Simon and Newell and the cognitive models of the Yale School, especially Carbonell's "Politics" program [1978, 1981] and the "Cyrus" program [Kolodner, 1980, 1983a], represent early, symbolically-based work in artificial intelligence with political relevance. Relatus follows in this vein.

While building from a distinct theoretical basis in AI, the Relatus project continues the research program of cognitive modeling begun by the Yale School. Our research program combines computer software and hardware advances [Rosenau, 1984] with theoretical developments in cognitive science, seeking to simulate human cognitive performance in politically relevant domains. Specifically, ours is an effort to model the symbolic structure of rational, and even irrational, thought [Abelson, 1958; Alker, 1979b; Simon, 1985].

Existing software technologies have largely proven inadequate for this purpose. However, guided by both their successes and their inadequacies we have developed Relatus. Relatus represents a
collection of new syntactic, lexical, referential, and representational technologies suited to cognitive modeling based on large texts. We can presently parse and represent several hundred pages of literal and explicit text at a rate of about forty pages per hour. Current limitations to strictly literal and explicit text makes the initial entry text tedious and frustrating. But further research in the areas of multiple word-sense processing, deliberative references, metaphor processing, and text coherence can lead to more fluent interaction.

In contrast to the top-down, domain-specific and knowledge-intensive approach of the Yale School, Relatus is grounded on a bottom up, domain-independent approach that relies primarily on syntactic knowledge. As a consequence, Relatus uses general methods to achieve a broad but shallow coverage, while the Yale approach (and most other contemporary AI approaches) rely on fairly deep but very narrow methods. In practice, it has not been possible to broaden these narrow methods. The Relatus strategy for surmounting this stumbling block is to build from a highly general base and gradually deepen the analysis across a broad front based on universally applicable methods rather than relying on ad hoc or domain-specific techniques.

1.2.3 The Coding and Meaning Problems

AI maintains a variety of techniques in its methodological armamentarium, such as logic programming and rule-based expert systems. Why should social scientists prefer natural language approaches to these simpler and more widespread methods? There are two major reasons, both of which revolve around the social basis of man in linguistic communities: the coding problem and the meaning problem.
The coding problem arises when natural language statements must be converted into a target formal language. Since formal languages such as logic and even rule systems are designed to yield single meanings unambiguously, the human coder must commit to a similarly narrow and precise interpretation of a text. This practice raises a difficulty. One particular interpretation may not suffice to capture the attributions signified to some political agents. Interpretation depends critically upon the situation and the knowledge of the interpreter. Thus, in the coding process, the coder imposes his or her own situation and knowledge upon the interpretation, rather than those of political agent or agents who are the focus of the modeling effort. This dilemma also plagues approaches which rely upon fixed-field codings for features.

The meaning problem is the inverse of the coding problem. Once a representation is available, the significance of statements in it may vary according to the situation of usage. Again, narrow and unambiguous representations in logical or rule languages provide no room for such variation. The committed formalist might respond that some logical formalism could capture the meaning variations across situations. This would be possible if and only if our coder could both recognize and predict all possible meanings within differing situations and the interpretive presuppositions of any political agent whose beliefs might be modeled. Of course, there may be arbitrarily many such situations, some of which may not be foreseeable. Also, the sheer number of such agents and the complexity of knowledge used by each to ascribe meaning preclude this approach. So, while translation to a
formal language uniform across situations and interpreters might sound plausible, it is in practice most unlikely to succeed for any non-trivial case.

For many social scientists, the meaning problem manifests itself as an empathy problem. The scientific observer places himself in the shoes of the subject of inquiry and attempts to simulate the subject's interpretive situation and knowledge. Of course, this sort of quasi-introspective approach is riddled with validity and reliability problems. This is where Relatus provides a systematic alternative, one that promises to help overcome these difficulties, as well as the twin problems of coding and meaning.

1.2.4 Reintroducing the Cognitive Factors of Political Explanation

The research interest that motivates Relatus recognizes that not all important social relationships are reducible to quantitative representation. For example, the beliefs and motivations of political actors -- surely critical explanatory factors in any political explanation -- cannot be reduced to quantitative indices without some degree of distortion. Faced with the non-reducibility of beliefs and motives, social scientific research practice has proceeded along one of two tracks: researchers either (a) quantify these factors anyway, ignoring the distortions thereby introduced; or (b) ignore belief and motivational factors altogether. By modeling political cognition in Relatus, we hope to include ideational factors in rigorous, replicable models.

Statistical and numerical techniques have proven to be quite useful tools for understanding the material, economic factors that
contribute to explanations of political outcomes. But, for the most part, the ideational factors remain, as Lasswell [1949: 47] put it, "qualitative, impressionistic, and conjectural." Certainly, psychologists have with some success applied statistical techniques, but political scientists rarely have access to the actors who make political decisions and cannot therefore administer intensive psychological instruments to these subjects. We must analyze instead the texts these actors produce — their letters, memoirs, journals, speeches, and propaganda.

We can infer little from these texts, however, by applying the techniques of computerized, quantitative content analysis [Stone, et al., 1966]. These techniques categorize words into classification schemes (sometimes derived through prior factor analyses) and apply inferential statistical techniques to the result. They can provide and have provided suggestive results at the macrosociological level [Namenwirth, 1973; Weber, 1978], at which only the very general thematic value concerns of texts need be studied. However, when the research interest involves the substantive inferences of particular political actors, such techniques are found to be lacking.

Computerized content-analytic procedures dissociate word-frequencies from the structural syntactic, semantic, and pragmatic relationships essential to language and to the meanings of particular texts. As a result, these techniques can provide summary information only of the value concerns embedded in a text. On the politically important issue of the direction of value orientations, these techniques are mute. For all levels of social scientific inquiry,
apart from the macrosociological level, existing computerized content analysis methods lack construct validity, and any results they produce at these levels are therefore suspect [Duffy, 1986a]. AI-based methods for analyzing the contents of texts, because they can incorporate the structural, linguistic relationships that inhere in texts, thus offer an alternative to the methodological limitations of computerized content analysis.

We seek to apply AI technology to all manner of politically relevant text. We construct textual models which incorporate the linguistic relationships which inhere in texts and without which texts lose their meanings. We plan to extend these models further, developing techniques which will support hypothesis formation and inference technologies for testing them. Relatus may be seen, therefore, as an extension of a research program begun by Lasswell, who sought throughout his career for methods with which we can incorporate ideational factors in our theory-based explanations of political phenomena [Lasswell, 1927: ix].

1.2.5 A Caveat

The recent spate of ballyhoo about AI in the popular press may tempt some to apply AI techniques too early to problems for which they are not yet well-suited. It is supremely important that one distinguish today's capabilities from tomorrow's, maintain a clear view of the current state of the technology, its contemporary limitations, and its near-term and long-term future possibilities.

In particular, we are skeptical of any attempt to directly apply today's AI techniques to the complexities of policymaking problems.
With Bloomfield [1986], we are even more skeptical of any reliance on existing AI techniques for guidance in foreign policymaking, where the complexities are most enormous. To be sure, applications of AI techniques may provide policymakers with a useful tool for modeling some aspects of complex policymaking problems, but only in concert with other modeling tools and in domains where the validity of program results has been affirmed. In the last analysis, it is most sensible for humans to remain the final arbiters of policy.

Contemporary hardware capabilities place strict constraints on the scope of AI capabilities, as do the algorithms implemented for that hardware. Current research in parallel processing promises to remove the Von Neumann bottleneck, which sharply limits the capabilities of today's serial processors, expanding by orders of magnitude the effective complexity which may be modeled and interpreted. In designing Relatus, we have kept this development in mind. Nevertheless, parallelism is no panacea for exponential algorithms, which render any problem intractable except perhaps in very small cases.

Contemporary storage devices and virtual memory machines also restrict the upper-bounds of the size of models that may be represented and analyzed in computers. This limitation may be lifted to some degree with the advent of video disc as a storage medium, lower costs for real memory, and concurrent processing machine configurations.

Another important limitation follows from the current state of research in natural language processing (NLP). Technology for NLP is presently in a pre-paradigm state. Although the Chomskyan and
derivative research programs dominate in the theory of syntax, lively competitors do exist. Moreover, many mutually incompatible theories compete in semantic theory, and pragmatic theory is only now beginning to receive the attention it deserves.

In sum, due to the contemporary condition of computer technology and the current state of linguistics, the glowing predictions of amazing capabilities sometimes foretold for AI will require much diligent scholarship as well as adequate funding targeted at the leading-edge in basic research. Although building computers with human-equivalent natural-language understanding remains an extremely difficult task, progress is being made and intermediate results will be available for social-science modelers.

1.3 Design Heuristics for Relatus

In 1980, when we sought out AI technologies to address the foregoing modeling considerations, we found that such technologies were neither available nor contemplated by mainstream AI researchers. We consequently took it upon ourselves to develop a system that would serve not only as a general environment for text analysis but also as an environment expressly suited to all social-scientific applications. While hermeneutics and lexical-interpretive semantics have guided our philosophic and linguistic strategies, methodological generality, representational fidelity, and computational tractability have guided our implementational tactics. We see the combination of these research heuristics as essential ingredients in any systematic effort to develop AI systems aimed at unconstrained domains, such as those native to the social sciences.
I.3.1 Hermeneutic Interpretation

Both our computational and social-scientific approaches are expressly hermeneutical [Mallery, Hurwitz, and Duffy, 1986]. We recognize, as a fundamental difference between the human sciences and the natural sciences, the fact that the objects of the human sciences — people — are themselves potential subjects of science. Unlike natural scientists, social scientists cannot simply impose an interpretative regime upon the people they study, no matter how broadly a consensus on that regime might be within the community of social science. As Habermas [1984: 13] put the case, "the conditions of validity of symbolic expressions refer to a background of knowledge shared by the communication community." The most pressing and most perplexing social problems are those in which this "background of knowledge" differs across social actors. Interpretation under differing histories and differing cognitive make-ups is a crucial prerequisite for sound social scientific practice, and so must also be a fundamental design criterion for any AI system intended to be useful for social-scientific modeling.

Strongly influenced by Tarskian truth semantics and logical empiricism, many AI practitioners have developed systems oriented implicitly toward an "objective" truth. While these approaches may address engineering concerns for many technical domains, they introduce distortions into textual interpretations by imposing a single interpretive regime where many alternative, and equally valid, interpretations are possible. These approaches have at best narrow
relevance for capturing the effects of ideology, culture, experience, and normative orientation.

These problems, which permeate social analyses, are best captured using methods which attempt to reproduce the life histories of social actors, deploying these histories to interpret texts and even to guide their reasoning and learning processes. While life histories sometimes supply interpretive presumptions found universally, or even across many social actors, they frequently provide interpretive presumptions peculiar to a specific group or even idiosyncratic. These prejudices [Gadamer, 1960] strongly influence the structure of belief, including even the ways in which social actors classify phenomena. The mix of information found in life-histories would be most difficult, if not impossible, to capture in ungrounded rules describing the set of concepts with which an idealized actor reasons. And even if this were possible, there would be no recourse when these rules "break down," [Winograd & Flores, 1986]. By having the life-histories available, it is possible not only to apply concepts and categories but also to learn them from experiential knowledge using induction and analogy.

Where such variations in the substratum of understanding exist, they evoke varying interpretations of the phenomena presented to the senses. This hermeneutic feature of understanding has long been familiar to humanists and most social scientists [Alker, 1978, 1979a], and now — thanks to Winograd and Flores [1986] — is being posed as a problem for AI. Hermeneutic aspects of understanding may yield to computational simulation. Relatus already captures certain referential abilities that depend on prior experience, and researchers continue to
develop informal reasoning abilities in the areas of induction and analogy. Metaphorical processing capabilities are somewhat more distant, but still on the horizon.

I.3.2 Lexical-Interpretive Semantics

Many social scientists are familiar with the conceptual dependency (CD) approach of Schank and Abelson [1977]. They may envision CD as covering the scope of AI approaches to textual analysis [Alker, 1975]. We have found that approach cognitively implausible and practically ineffective. We have consequently developed an entirely different approach which we believe is both more plausible and more effective.

Schank and Abelson decompose lexical terms appearing in a text to a fixed set of "semantic" or "conceptual" primitives. Our approach — lexical-interpretive semantics [Mallery and Duffy, 1986] — is not decompositional. We refrain from the reduction of lexical terms to primitives for five reasons.

1. No principled grounds exist for the justification of linguistic primitives [Mervis and Rosch, 1981: 104-106].
2. Reduction to a fixed stock of concepts inevitably distorts the meanings of terms and thus also of textual contents.
3. CD provides no room for an account of linguistic innovation and evolution.
4. CD imposes a single, putatively "objective", interpretation upon textual contents, precluding the careful study of interpretation and classification according to the interpretive horizons of any life-world other than those of the CD designers.
5. The available experimental evidence supports a lexically-based model of memory [Fodor et al., 1975; Hayes-Roth and Hayes-Roth, 1977; Fodor et al., 1980; Gentner & Landers, 1985].
In lexical-interpretive semantics, the lexical items in the text (after canonicalization for the removal of syntactic, but not semantic, variation) are themselves the representational designators. These lexical terms are interpreted on the basis of (a) the system's general knowledge of the token and its usage, and (b) the discourse context, including the intentions of the text producer and the intended audience of the text [Grosz & Sidner, 1986].

While the decompositional approach finds equivalent meanings through reduction to common conceptual primitives, lexical-interpretive semantics finds these equivalences by dynamically creating meaning-congruence classes for lexical realizations. A meaning congruence is a set of intentionally equivalent but lexically alternative ways of expressing the same propositional content. Because determining these congruence classes is itself an interpretive process, lexical-interpretive semantics may be seen as a early level of hermeneutic text processing [Mallery, Hurwitz, and Duffy, 1986].

1.3.3 Generalizability, Fidelity, and Efficiency

Three general design heuristics — generalizability, fidelity, and efficiency — have guided our implementational work from the start. These, in turn, justify our adoption of a non-decompositional, lexical-interpretive approach.

1.3.3.1 Generalizability

We envision Relatus as a general text analyzer, capable of performing analyses at all linguistic levels, and applicable across the broad range of complex problems of interest to social scientists. Thus,
we would like Relatus to be able to process and understand a wide variety of texts incorporating an essentially unlimited variety of concepts and terms.

Text understanders based upon the CD approach [e.g., Dyer, 1983] require elaborate hand-codings of each possible lexical term appearing in a text, translating each into a different composition of semantic primitives for each sense in which the term might be used. This process would consume an inordinate amount of time and is thus unfeasible for serious social-scientific applications.

Lexical-interpretive semantics, on the other hand, captures the meanings of terms by representing context-specific textual definitions and textual examples of their usages (relative to contexts) in the same (semantic network) form as all other texts are represented. When texts are subsequently processed, this information is used to determine meaning congruences at interpretation time.

To clarify, suppose that, instead of reducing lexical items to conceptual primitives, a CD text understanding program simply represented the term in a semantic network, relating it to other terms in the text by means of associational links. Now imagine that the system then created associational links between the term and its conceptual primitive. This, in essence, would be how a meaning congruence class would be related to a lexical term. However, instead of determining the appropriate meaning congruence class a priori, as in CD approaches, we would instead determine it a posteriori, as a function of the lexical relationships already represented in the semantic network which constitute a model of the life-world of the
interpreter. Likewise, the terms which count as "primitives" (in the CD sense) or as signifiers of meaning congruence classes (in our sense) are determined within the context of the life-world of the interpreter, not imposed upon the interpreter from without.

In our view, conventional, socially-shared meanings emerge among linguistic communities over time. As individuals interact linguistically in these communities, they educe conventional meanings. While many variations may arise, all competent native speakers will naturally share a large number of meanings and forms of expression. Understanding these differences and similarities becomes a crucial task for the social scientist. As an important consequence for natural language modeling, the repertoire of conventional meanings captured by a text understanding system increases as experience with different linguistic communities and text corpora broadens. Thus, as methodological generality contributes to the ability to parse and represent large and variegated texts, the ability to capture the conventional meanings of a variety of political genres becomes conceivable.

1.3.3.2 Fidelity

Relatus maintains a high degree of fidelity. It avoids systematic distortions of textual contents in two ways — through a hermeneutic interpretive approach and by maintaining full information models.

Relatus assumes that no single, rigid interpretive paradigm can provide interpretations valid across time and across interpreters. Instead, Relatus interprets text through the interpretive template constituted by the system's effective history, the prior history of
interpretations [Gadamer, 1980]. Because texts are thereby interpreted according to the life-history of the interpreter, interpretations in Relatus are hermeneutic, and thus more appropriate for social-scientific concept and theory formation [Schutz, 1967] than are anti-hermeneutic, decompositional approaches.

By definition, a model summarizes some phenomenon. Modelers attempt to summarize phenomena without distortion, reducing them to their essentials and representing only those essentials in models. However, when we model a text, we are recovering from a linearization generated by the text producer another model — one based upon the text producer's own determinations of the essential aspects or features of the subject-matter of the text.

One must assume that the producer of an historical narrative, for instance, has already provided in the text a model of the course of events which has been streamlined to its essentials (as perceived by the text producer). There is thus nothing to gain and much to lose by reducing the text further. For this reason, the design of Relatus incorporates a full information criterion. Textual contents are mapped to semantic network representations with zero information loss.

Reduction to semantic primitives, however, results in the loss of much information. For example, in the conceptual dependency paradigm, verbal predicates like "migrate", "flee", "defect", and "vacation" would each be reduced to an identical primitive. As a result, the substantive, politically relevant differences between the meanings of each term would be lost. Not only is this cognitively implausible [Fodor et al., 1975; Hayes-Roth and Hayes-Roth, 1977; Fodor et al.,
1980; Gentner & Landers, 1985], but it does severe damage to the validity of any social scientific results derived therefrom.

I.3.3.3 Efficiency

The size of the largest text that may be modeled and interpreted tractably increases as the overall efficiency of the system increases. The inherent complexity of social and political phenomena therefore demands that the implementation operate as efficiently as possible. Efficiency is particularly important for central modules, such as parsing and reference. If inefficient, the range of tractable reasoning and learning operations would diminish. The efficiency of reference is particularly important in this regard.

Relatus is designed for efficient text processing. Mappings to semantic structures are performed from the syntactically canonical output of the Relatus Parser. Syntactic canonicalizations are information-preserving and greatly enhance the efficiency of reference [Mallery and Duffy, 1986] -- the process of referring to configurations of related lexical terms in a semantic network. Reference is a key operation for any AI natural language system. In Relatus, the time-complexity of reference operations is independent of the size of the semantic network [Mallery, 1986]. Thus, apart from hardware limitations, as the size of the text modeled increases, no inherent degradation in the performance of referential operations need result.

Appreciation of the performance characteristics of the Relatus reference system requires substantial technical discussion which is beyond the scope of this paper. Suffice it to say that, to the best of our present knowledge, this is the best subgraph isomorphism algorithm
suitable for use in a system based upon lexical-interpretive semantics.

I.3.4 Unconstrained Domains

AI program developers generally worked within very limited "micro-worlds." When producing practical programs, they generally constrain the program to operate only within a specific, well-defined problem space. When producing programs to demonstrate aspects of a cognitive theory, they ignore (for very defensible, practical reasons) problems that do not pertain directly to their scientific point. As a result, most, if not all, existing AI programs simply will not scale up to process large texts in a manner which is both comprehensive enough and of sufficient inferential power to satisfy the demanding requirements of the social scientist.

Only when this goal figures in design considerations from the inception of the research will any AI program be able to meet these requirements. We have directed our research toward the goal of producing a system which can serve for the social-scientific analysis of textual data a role analogous to that which systems like SPSS and SAS serve for the social-scientific analysis of numerical data. Our experiences with other AI programs taught us that most existing AI systems can never scale up. If we were to meet our goal, we had to start from scratch, grounding Relatus on the design principles outlined above.

---

57 Readers interested in further details should consult Mallery and Duffy [1986] and Mallery [1986].
Strict adherence to these principles has in turn allowed a system to emerge which we believe will have many practical social-scientific applications. Before discussing possible applications, the current state of the implementation should be described.

I.4 Stages of System Development

Relatus will develop in two stages. The two stages correspond to two broad classes of reference, immediate reference and deliberative reference. Immediate reference merges lexical items into a syntactically canonical semantic network by "reading off" the data structures without any significant reasoning. Deliberative reference is more complex. Building upon immediate reference as its base, deliberative reference may involve logical reasoning, learning operations including analogy and metaphor, and determinations of meaning congruence classes.

These two categories of reference are distinguished by the need to track referential dependencies. The audit trails of the dependencies underlying particular deliberative references are necessary in case information on which a reference depends later becomes invalidated. This could require retraction of the reference. In contrast, immediate references do not require dependency tracking [Mallery & Duffy, 1986; Mallery, 1986].

At this writing, we have successfully completed the first stage and are proceeding with the second. The two stages are:

1. **Stage One** -- Literal and explicit sentences, which exploit only single senses of words, are referenced incrementally from a text. This process yields a referentially integrated semantic graph model of the text suitable for reasoning
operations and question answering. Users can query these models with literal and explicit questions. Since deliberative reference is not yet available, reasoning cannot yet help maintain the coherence of a textual interpretation. Textual coherence must therefore be explicit in the text. The significance of this stage is the ability to construct text models suitable for reasoning and learning operations, a necessary precursor to the second stage.

2. **Stage Two** — Deliberative reference augments immediate reference, resulting in more intelligent system behavior. While immediate reference procedures may not find a particular literal referent, deductive, inductive, analogical or even metaphorical inference procedures may yield a non-literal referent that satisfies relevant referential constraints. But speakers frequently rely on the ability of hearers to predict the desires and plans of intentional agents. Thus, deliberation must ultimately simulate this, resolving certain kinds of references. The open-ended nature of deliberation demands full reasoning, planning, and learning capabilities to effectively handle unrestricted text. Deliberative models can interpret input and generate reasoned responses to queries based on the information they contain.

I.5 Substantive Goals of Relatus

I.5.1 Relatus as a Social Science Workbench

Alker [1975] considers artificial intelligence to be a descriptive technique. Yet to maximize its usefulness for social–scientific theory construction, we would like Relatus to incorporate operations to support hypothesis testing [Thorson and Sylvan, 1982; Chan and Sylvan, 1984]. Inferential statistical techniques also allow us to make counterfactual statements to see what conditions would then obtain. We can ask "what if" questions. Given a particular parameter estimation, for example, we can assess the impact a change in an independent variable would have upon a dependent variable. We would like to be able to perform analogous operations with textual data in Relatus, making it a tool for social–scientific inference.
By manipulating conditions counterfactually, Relatus will provide a computational framework in which to test hypotheses concerning substantive problem domains. As noted above, all interpretation in Relatus is conditioned by a background of previous interpretations. The prior interpretations, which guide subsequent interpretations, can be manipulated systematically, affording an opportunity to answer counterfactual ("what if") questions.

For example, given that an interpreter deduces a proposition from a universe of prior beliefs, we could assess the importance of any particular belief for the deduction by removing it or changing its substantive contents. Likewise, should we wish to know the deductive consequences that might emerge if a text-interpreter believed a particular proposition, we could add the proposition to the interpreter's universe of beliefs, comparing the deductions that emerge with the proposition to those that emerge without it.

But hypothesis testing does not exhaust the scope of possibilities that lie ahead. Given that we are developing an environment for cognitive modeling, we hope to be able to simulate analogical and metaphorical reasoning, situation-based planning, inductive concept formation, and even a situation-based form of deductive inference from incompletely-specified information.

These sorts of capabilities will allow social scientists to specify the historical experience of social actors and perhaps supply some conceptual structures directly. By doing this for a series of political agents and by describing the environment in which they are
embedded, simulations of purposeful, political behavior become possible.

I.5.2 Substantive Applications

To make our goals more concrete, it is helpful to outline briefly a few of the substantive applications which we envision for Relatus. Among these applications are the following:

I.5.2.1 Historical Analysis

Relatus may model an historical account of a situation or a particular actor's experience. Precedent-based reasoning [Alker & Christensen, 1972; Alker, Bennett & Mefford, 1980; Winston, 1984; Mefford, 1986] would be used to analyze the current situation in terms of previous situations. In fact, this technique can be applied recursively, as Winston [1984] shows, to synthesize novel precedents by combining pieces from various past situations.

This approach seems very promising for understanding complex decision-making processes, such as those found in international conflict and foreign-policy decision-making. It provides a principled way of simplifying, with a plausible paraphrase, the overwhelming complexity that actually goes into these decisions. In fact, arguments from complexity theory, "bounded rationality" and circumscription [McCarthy, 1980a, 1980b; Williams, 1985] suggest that rational agents, including those in bureaucratic organizations, cannot consider the full range of possible actions and consequences because deductive operations explode (combinatorically) too quickly.
Analogical and precedential reasoning provide an both an account and a mechanism for simulating important reasoning processes that can account for behavior and which are computationally tractable. They are tractable because the search space can be small, being limited to "similar situations," and they are also logically incomplete. But, as historical experience increases, the likelihood of finding a similar situation and an adequate action increases. Thus, these methods may be good ways to paraphrase the inferential behavior of political actors both in computational simulation and in qualitative analysis.

1.5.5.2 Theory-Based Reasoning

Another analytical approach would represent a theory and apply it to individual cases. A propositional inventory of a particular social theory is first input in English. Textual accounts of individual cases are then input. Relatus is next asked questions about the cases. Its answers are theory-based interpretations of the story. The predications of different theories about different cases can then be compared. The inferences made by each theory may then be contrasted to summarize the differing substantive implications of the theories.

A variant of this approach would focus the analysis on the theoretical accounts to estimate their coverage, internal structure, logical consistency, and completeness. Propositions may also be added to or deleted from a particular theory to see whether these make any difference in the interpretations of and inferences about the narrative. Work is presently underway to represent and use a Lateral Pressure theory [Choucri & North, 1974] in this way.
1.5.2.3 Theory Formation

Given algorithms for inductive and analogical learning, it becomes possible to learn theories from empirical narratives. For example, we have recently modeled sequential prisoner's dilemma game-protocols [Hurwitz, Mallery, Alker, Duffy, 1986]. These protocols are empirical narratives. They contain transitions between various phases. Of particular interest to the peace researcher are the transitions to community (mutual cooperation) and anarchy (mutual defection). We have been working on new inductive learning algorithms that will be able to identify regularities in the beliefs, intentions, and expectations of players that leads to the phase transition. By examining a larger number of cases, it will become possible to identify the range of player reports preceding these transitions. This constitutes learning a micro theory.\textsuperscript{5858} While this is only a first, modest essay, it does suggest the prospect of applying these kinds of learning techniques [Dietterich & Michalski, 1983] to develop classificatory theories.

1.5.2.4 Social Network Analysis

Relatus may be used to model social networks. Relatus converts textual data to networks of relations. Textual inputs of social relations can construct large-scale textual models of social networks. Social network analysts typically want to analyze more than one type of social relationship at a time, finding the similarities and differences between large chunks of network structure, and perhaps inducing a

\textsuperscript{58}Unfortunately, this work remains incomplete and must be reported later.
typology of objects (persons and groups) and relations in a social network.

Traditional graph-theoretic formalisms cannot tractably perform these multiple-relationship sorts of analysis, and have thus proven less than useful for modeling large-scale social systems. For this reason, cluster analysis and multi-dimensional scaling are often applied as data reduction techniques for use in describing large networks. These techniques are problematic, since they weight each relationship equally, introducing systematic distortion into any subsequent analysis. One can adopt a strategy that weights the links differentially, but it is difficult to see how a demonstrably valid set of weights can be formulated. 59

Since the time-complexity of reference in Relatus is independent of the size of the database, Relatus will afford social scientists the opportunity to analyze much larger networks of social relations than have heretofore been possible. Such analyses need not rely on the potentially distorting effects of cluster analysis and multi-dimensional scaling, noted above, because no data reduction techniques will be needed. Instead, multiple-relation analyses may proceed using English-language descriptors of the relationships. Since no scaling or clustering need be performed, there would be no need to provide numerical weights on the relation descriptors.

I.5.2.5 Intentional Game-Theoretic Analysis

59 For a discussion of these issues, see [Laumann, 1979].
Relatus can improve game-theoretic models. Political scientists often construct such models for understanding negotiations and conflicts. The purely mathematical nature of those models, however, force researchers to ignore the important intentional aspects of any player's game play, thereby affecting negatively the validity of game-theoretic results. Relatus has been used to model protocols of sequential prisoner's dilemma games [Alker, Duffy, and Mallery, 1985; Hurwitz, Mallery, Alker, Duffy, 1986]. Relatus can include those aspects by incorporating players' responses to protocols designed to elicit their intentional attitudes and attributions during the conduct of the game. It can also recognize phases of play, and thus, classify types of play.

1.5.2.6 Political Belief Analysis

Relatus can help us improve our understanding of the nature of political beliefs. Studies of political belief have thus far been either large-sample correlational summaries of a narrow range of issue positions [e.g., Converse, 1964] or anecdotal interviews with individual respondents [e.g., Lane, 1962]. We would like to know much more about political beliefs, inferencing, and political ideology than can be gleaned from issue correlation studies, and we would like to do so in a more rigorous manner than is provided by a single interpreter's account of particular interviews.

Relatus will allow the examination of the political beliefs and inferences of individual actors in a rigorous and replicable manner, affording the opportunity to test hypotheses concerning the addition, deletion, or modification of beliefs. For example, suppose a
respondent believes a large set of political propositions, some of them dependent upon one another and others contradictory with one another. Suppose further that the respondent became convinced of the fact that two such propositions were contradictory and opted to discard one of them. Under what conditions will the respondent also discard propositions dependent upon the belief in the earlier proposition, and on what conditions will he retain the belief, searching for additional support for it? Relatus can be extremely useful in this sort of research, particularly when the web of beliefs is quite large and quite dense.

Applying the technique of theory-based reasoning, described above, one might model interviews with particular respondents, interpreting them in the context of a variety of models of theories about belief, such as those of cognitive consistency theory, cognitive dissonance theory, cognitive balance theory, attribution theory, etc. Each theory might suggest hypotheses about particular respondents, given the character of their stated beliefs. By interpreting respondents' statements in the context of those beliefs, such hypotheses may be drawn out of Relatus, and they can subsequently be tested.

I.5.2.7 Cognitive Analysis

Relatus implements a new perceptual theory based on constraint posting [Mallery and Duffy, 1986]. The theory is consistent with various experiemental psycholinguistic findings. [Fodor et al., 1975; Hayes-Roth and Hayes-Roth, 1977; Fodor et al., 1980; Gentner & Landers, 1985]. For this reason, Relatus promises to become a useful test-bed for simulations of human linguistic performance. Other simulations are
also possible. For example, some experiments were conducted during 1984 in which visual images derived from physical sources were mapped into the semantic representation in Relatus using the reference system, 6060

1.6 The State of the Implementation

Most components of Relatus are implemented as message-passing agents. This means that data structures in Relatus are instantiations of conceptual object types. These instantiations pass messages to one another. When an instantiated agent receives a message, it invokes its handler for that message. Since different agent types may have different handlers for the same message, agents which send messages need not reason about the type of agent receiving the message. This technique reduces the need for complex conditional expressions in the code, resulting in a significantly simplified control structure [Hewitt, 1977].

Message-passing techniques also afford the opportunity to group operations together by function, enhancing the debuggability of the system. Additionally, since grouped message-handlers constitute protocols for communication between program modules, the modules themselves may be redefined without the need for massive surgery on the source code. Various implemented components of the present system are described briefly below.

1.6.1 Knowledge Representation

60Some of the vision research is briefly mentioned in [Mallery, 1986]. The prospect of embedding a parser inside a semantic representation is discussed in [Duffy, 1986b].
Relatus Belief Systems are central components around which all processing revolves. They are central because they contain the knowledge. The syntactic and reference components aim to build correct semantic structures in belief systems. Systems for reasoning, which are presently under development, use the semantic structures as both input and output. Belief systems hold and manage a special class of graph structure formed by labeled binary relations. This and related classes of graph structures are frequently known as semantic networks because they are used to represent information derived from natural language texts. Each node in the graph structure is a message-passing agent that supports a variety of operations, such as semantic inversion (discussed below).

Belief systems perform many operations. The major operations include syntactic analysis using the Relatus parser, semantic reference using the constraint-interpreting reference system, and question answering. More minor operations include tracking the source sentences for semantic structures, displaying knowledge structures, dumping to file servers, processing user-defined directives (actions for the system to take when given imperative sentences by the user), and an "if-added demon" facility (which invokes a pre-defined procedure whenever a particular token-type is added). Unlike conventional knowledge representation schemes, Relatus belief systems are instances of message passing agents. This makes it easy to create new belief systems and to put similar or different texts in them, and in turn, provides a means for simultaneously simulating different interpretive regimes.
I.6.2 Syntactic Analysis

The Relatus Parser is a transformational, deterministic English syntactic transducer with unbounded sentential look-ahead. The parser produces a directed cyclic graph that describes the syntactic contents of a sentence. Interwoven within this cyclic graph are directed acyclic graphs that describe the syntactic deep structure and syntactic surface structure of a sentence. All non-terminal nodes in the parse graph are implemented as message-passing agents. These agents themselves perform the analysis of a sentence and its components.

During analysis, control is progressively passed through the graph to various clausal and phrasal agents, each of which supervise analysis within their scope. The resulting parse tree scopes embedded constituents by their categorial role in their parental node. These agents also supervise the process of constraint-posting (discussed below).

Components are included within the parser for lexical insertions of ellipses, punctuation handling, possessive handling, English-Arabic numeral translation, and intrasentential anaphora resolution. Clausal and phrasal agents can also reference syntactic structures in a belief system, so that the latter may reason over and learn syntactic knowledge. The parser is discussed more fully in [Duffy, 1986b].

Because sentential look-ahead is unbounded, the parser is not deterministic in the sense of Marcus [1980]. However, because operations terminate at clausal boundaries, the parser is effectively deterministic. It remains a polynomial LR(k,t) algorithm, although k is variable, not constant. See [Berwick and Weinberg, 1984: 192] for a discussion of the time complexity of LR(k,t) parsers.
I.6.3 Categorial Disambiguation

A Categorial Disambiguator disambiguates the categorial roles (parts of speech) of word-tokens. Relatus includes a facility for disambiguating the categorial roles that words play in particular sentences. The categorial disambiguator is based on a constraint propagation scheme [Waltz, 1975]. When an ambiguous category is encountered, the disambiguator exploits the constraints propagated from its neighbors. The successful disambiguation of one ambiguity in a sentence propagates additional constraint which may be helpful in disambiguating other categorial ambiguities in the sentence. Categorial disambiguation is an important, computationally inexpensive prelude to more expensive semantic disambiguation procedures involved in parsing and reference [Duffy, 1986c].

I.6.4 Constraint-Interpreting Reference

A Constraint-Interpreting Reference System is an interpreter and an extensible set of declarative constraints that are used to find and build graph structure in the semantic representation [Mallery, 1986]. The constraints comprise a language for describing graph structures. Constraints are bundled in units called reference specifications. These are the actual structures processed by the reference system. An important feature of this system is that, unlike many earlier algorithms, it can reference subgraphs in time independent of database size.

The reference system relies on the self-indexation of belief systems to retrieve graph structures. Belief systems index themselves
based on a theory of particulars involving token-typing and bi-directional relations. In contrast to retrieval systems found in systems which classify decomposed conceptual primitives, the self-indexation and reference procedures of Relatus do not require extensive, fallible high-level analyses. This bottom up approach characterizes a general approach in the Relatus project: Concentrate on general but shallow natural language processing. Building from a broad and general foundation gradually extend the depth of analysis in those areas that will yield the highest payoffs soonest (i.e., whose generality solves the most textual problems).

1.6.5 Sentential Constraint-Posting

A Sentential Constraint Poster maps parse graphs into belief systems. This process utilizes both surface and deep structure from the syntactic analysis. While noun-phrase quantification [Mallery, 1985] is analyzed at surface structure, deep-structure grammatical relations are used to construct sentential reference specifications. This process is called sentential constraint-posting because the various piece of information comprising the sentential description are incrementally collected in a single sentential reference specification. The sentential reference specification is referenced in a belief system only when complete.

Sentential constraint-posting may be contrasted with earlier approaches, which performed incremental references before building complete sentential descriptions. Using both deep structure and constraint posting minimizes unnecessary backtracking in both constraint posting and sentential reference [Mallery and Duffy, 1986;
Mallery, 1986]. By successive referencing of sentences in a text, Relatus performs intersentential reference — construction of a belief system graph structure in which references to the same entities across sentences are maintained in the semantic representation. Thus, the sentential constraint poster is a key component in the ability to construct text models consistently.

I.6.6 Semantic Inversion

The Semantic Inverter implements the inverse of constraint-interpreting reference. Given a node in the semantic graph structure and some constraints for the incorporation of other nodes, the semantic inverter traverses the semantic graph structure by following all relations from the initial node that satisfy the incorporation constraints. As it performs this traversal, it constructs reference specifications describing the graph traversed. These reference specifications can then be referenced using the reference system. This makes possible automatic construction of reference specifications tailored to particular tasks (e.g., question answering, inductive generalization). Semantic inversion makes possible conversions to and from declarations (semantic graph structures) and procedures (reference specifications).

I.6.7 Question Answering

A Question Answering Facility can answer literal and explicit questions about the knowledge represented in belief systems. The infrastructure provided by the parser, the reference system, and the
semantic inverter make the algorithm that actually answers questions
straight-forward and general.

In most cases, the problem of answering questions is reduced to
(a) syntactic transformation of an interrogative to a declarative form,
(b) construction of a belief-system graph structure for the question
which is isolated from the rest of the belief system, and (c)
construction of a reference specification which, when referenced in the
belief system, yields the correct answer. In questions where there are
unknowns (e.g. "Who was wanted by the secret police?"), the unknown in
the declarativized query (e.g., 'x' in "The secret police wanted x.")
is found and the semantic inverter builds a reference specification for
the question from the perspective of the unknown (i.e., the reference
specification describes an x such that the secret police wanted it).
This specification is then referenced in a belief system to yield the
answer.

Like the other components of Relatus, the power of the
question-answering facilities rest not in the depth of its processing
but in the breadth of the questions it can handle. For any general
syntactic form that it handles (and that earlier processing levels
handle), it can answer literal and explicit questions. This is a
marked contrast to earlier AI approaches, in which the capacity of the
implementor to write unlimited numbers of "if-added demons" severely
limits the scope of questions which may be answered.

1.6.8 Lexical Knowledge

The Relatus Lexicon provides lexical constraints to the parsing
and reference components. This 16,000-token lexicon stores the case
and complement features of nouns and verbs as well as verb conjugation information. It also points to the synonyms and antonyms of words, and to their various alternate categorical forms. This lexical knowledge parameterizes operations within the parsing and reference components.

An new lexicon is currently under development using an experimental generic lexicon shell. The shell allows system developers to modify or extend the fields (subcategories) of words and idiomatic phrases so that lexical parameters necessary but not anticipated prior to the actual instantiation of the lexicon. The lexicon shell is also language-independent and exploits certain data compaction strategies.6262

I.6.8.1 Lexicon Editing

A Lexicon Editor facilitates user extensions to the lexicon. Users can define the technical senses of certain words, add words that are not in everyday usage, and declare lexical constraints and subcategories. The editor is menu-driven. This means that the user's options are displayed in a pop-up window, or 'menu', from which choices can rapidly be made (using a mouse) without unnecessary typing. To facilitate use of the system, the lexicon editor is invoked automatically during a parse, whenever information about a particular word is needed. The lexicon editor constitutes a structured method for the acquisition of lexical information from human users.

I.6.8.2 Lexicon Patching

---

62A detailed description of the lexicon shell is provided in [Duffy, 1986d].
A Lexicon Patch Facility remembers users' lexicon changes and saves them in textual files. During subsequent sessions with Relatus, these 'patch files' can be automatically reloaded into the lexicon. Routines have been written for performing various operations (e.g., merging) on patch files. All patches pass through a series of filters which detect any unwarranted structures which might otherwise confuse parsing or reference procedures. These filters guarantee the integrity of the lexicon.

I.7 Assessment of Relatus as a Research Environment

Relatus represents a significant foundation for textual modeling. Code abstraction practices and object-oriented programming techniques have enhanced the system's flexibility, modularity, extensibility, portability, and understandability.

Because natural language is complex, and because many modules are required to support an acceptable research environment, the system has grown quite large. To counter creeping complexity and to enhance productivity, we have designed and implemented debugging tools, user interfaces, and software organization methods which enable researchers to quickly diagnose and correct bugs. These include the following:

1. Systems for organizing procedures for easy summary and location.

2. User interfaces for parsing and representing text.

3. The menu-driven lexicon editor (discussed above).

4. A Relatus mode for the machine's text editor, making interaction with system more intuitive for users.

5. Mouse-sensitive inspectors for browsing through the belief systems and parse graphs.

7. Adaptation of the machine's mail facilities to generate the relevant details in bug reports.

8. A background bug reporting facility, which allows the system to generate its own bug reports automatically under certain prespecified conditions.

These features have enabled us to maintain control over the development of the system, and make further developments possible.

I.8 Relatus as a Tool for Multi-Disciplinary Scientific Progress

Two major research divisions emerge in the Relatus project. On the one hand, the state of the Relatus implementation already allows research in computational topics such as natural language processing, automatic inference, and automatic programming. On the other hand, the extent of implementation of these AI techniques enables classes of social scientific research based on them. But understanding intelligence, which, for our purposes, we take to be the processing structure of a rational agent capable of linguistic communication, requires cross-pollination between the social sciences, the cognitive sciences, and especially artificial intelligence.

Combination of these somewhat diverse research emphases in a single computational environment promises to provide not just a focal point to bring together researchers from various domains, but also a means of concrete theory-testing which can cumulatively consolidate progress in each. Progress in one area is immediately made available to others by implementation. Researchers in another area may then make advances in their own domains building those of other areas, and, in turn, feed back them back into the system. In this way, a system like
Relatus can become a vehicle for communicating scientific results, advancing many research areas simultaneously.

I.9 Acknowledgments

This paper was improved by comments from Hayward R. Alker, Jr. and Roger Hurwitz. The authors are responsible for the content.

This report describes research performed at the Artificial Intelligence Laboratory of the Massachusetts Institute of Technology. Support for the Laboratory's artificial intelligence research is provided in part by the Advanced Research Projects Agency of the Department of Defense under Office of Naval Research contract number N00014-80-C-0505.

I.10 References


Bonham, Mathew, and Michael W. Shapiro, eds., Thought and Action in Foreign Policy, Basel: Birkhauser Verlag, 1975.


Duffy, Gavan, "The Validity Problem in Content Analysis", mimeo, The University of Texas at Austin, Department of Government, 1986a.


Ricoeur, Paul, "The Model of Text: Meaningful Action Considered as


APPENDIX II

Temporal Indexation Code

;; Range searching using self-balancing 2D trees.

;; Node definition.

(defun print-2d-node
  (node stream &optional prinlevel &rest ignore)
  "Print function for 2D nodes."
  prinlevel ;ignore
  (when node
    (format stream "[\"S\"\"]
                (2d-node-x node)
                (2d-node-y node))))

(defstruct (2d-node (:predicate 2d-node-p)
                     (:print-function print-2d-node)
                     :named)
  "A node for a two-dimensional tree."
  (x nil :type fixnum) ;x value
  (y nil :type fixnum) ;y value
  (parent nil :type 2d-node) ;node above
  (left nil :type 2d-node) ;lower in dimension
  (right nil :type 2d-node) ;higher in dimension
  (depth 0. :type fixnum) ;depth in tree
  (depth-right 0. :type fixnum) ;max right
  (depth-left 0. :type fixnum) ;max left
  (lowest-below nil :type 2d-node); in opposite dimension
  (highest-below nil :type 2d-node)); in opposite dimension

;; Balance Substs

(defun balance-degree (node)
  "Returns the absolute value of the difference between a NODE's
   left-depth and right depth."
  (abs (- (2d-node-depth-right node) (2d-node-depth-left node))))

(defun 2d-node-balanced-p (node)
  "Non-null if NODE is in balance."
  (> 2. (balance-degree node)))
Storing a node

Top-level form is STORE-2D-NODE.
To support intern-time balancing, each node remembers the
nodes with the highest and lowest values in the opposite
dimension.

(defun 2d-node-equal (node1 node2)
  "Non null when two nodes have the same x and y values."
  (and (= (2d-node-x node1) (2d-node-x node2))
       (= (2d-node-y node1) (2d-node-y node2))))

(defun maybe-remember-node-below (below node depth position)
  "Compares a node to be interned BELOW with NODE's encached
  lowers. If node is lower than or higher than the encached
  node for a particular dimension, it replaces the encached
  node. If there is no encached node, BELOW is encached."
  (let ((already-lowest-below (2d-node-lowest-below node))
        (already-highest-below (2d-node-highest-below node)))
    (cond ((and (null already-lowest-below)
                (null already-highest-below))
          (setf (2d-node-highest-below node) below)
          (setf (2d-node-lowest-below node) below))
          (position
           (let* ((accessor (if (oddp depth)
                               #'2d-node-x
                               #'2d-node-y))
                   (low-below (funcall accessor already-lowest-below))
                   (high-below (funcall accessor already-highest-below))
                   (value (funcall accessor below)))
            ;; use >= and <= to ensure that if there are two nodes
            ;; with the same values in a dimension, the most recent
            ;; insertion falls to the bottom of the tree.
            ;; This reduces the need for rebalancing replacement
            ;; nodes (see 2d-balance-1).
            (when (and (>= value high-below)
                        (not (2d-node-equal below already-highest-below)))
              (setf (2d-node-highest-below node) below))
            (when (and (<= value low-below)
                        (not (2d-node-equal below already-lowest-below)))
              (setf (2d-node-lowest-below node) below))))))
(defmacro .store-node-here. ()
  "We've run out of tree, so store the node here.
  Expands inside STORE-2D-NODE."
  `(progn (setf (2d-node-depth node) depth)
           (when parent (2d-node-attach node parent direction)))
)

(defmacro .store-node-left. ()
  "The predicate succeeds in the appropriate dimension,
  so follow the left branch. Expands inside STORE-2D-NODE."
  `(progn (maybe-remember-node-below node tree depth position)
           (store-2d-node-1 node (2d-node-left tree) tree
            (+ 1. depth) :left node-x node-y)))

(defmacro .store-node-right. ()
  "The predicate fails in the appropriate dimension,
  so follow the right branch. Expands inside STORE-2D-NODE."
  `(progn (maybe-remember-node-below node tree depth position)
           (store-2d-node-1 node (2d-node-right tree) tree
            (+ 1. depth) :right node-x node-y)))

(defun store-2d-node (node tree)
  "Store NODE at DEPTH in TREE with PARENT."
  (store-2d-node-1 node tree)
  ;; the root may have changed.
  (if tree (2d-node-root tree) node))
(defun store-2d-node-1
  (node tree &optional parent (depth 1.) direction
    (node-x (2d-node-x node))
    (node-y (2d-node-y node))
    (balance-p t)
    from-balance-direction)

"Does the work for STORE-2D-NODE."
(cond ((null tree)
      ;;ran out of tree, so attach node.
      (.store-node-here.)
      ;; Otherwise, apply the predicate.
      (t (let* ((tree-x (2d-node-x tree))
               (tree-y (2d-node-y tree))
               (position
                (when parent
                  (position-in-parent tree parent t))))
               (y-p (oddp depth))
               (nvalue (if y-p node-y node-x))
               (tvalue (if y-p tree-y tree-x)))
      (cond ((and (= node-x tree-x)
                  (= node-y tree-y))
            ;;already in tree, so return non-null.
            (nvalue tvalue)
            ;;go left if less
            (.store-node-left.)
            (tvalue (if nvalue tvalue)
                    ;;go right if more
                    (.store-node-right.)
                    ;;otherwise they're equal, so
                    ;;choose a direction that minimizes
                    ;;later balances.
                    (or (> (2d-node-depth-right tree)
                          (2d-node-depth-left tree))
                        (= :right from-balance-direction)
                        (.store-node-left.)
                        (t (.store-node-right.)))))))
      ;;Do all this on the way out of the recursion.
      (cond (tree (adjust-depths tree)
                 (cond ((and balance-p
                           (not (2d-node-balanced-p tree)))
                        (2d-balance tree))
                        (t tree)))
        (t node)))

;; Code for finding the appropriate replacement for an
;; imbalanced node.

(defun find-replacement (node y-p direction)
  "Finds a balance replacement for NODE."
  (multiple-value-bind (child comparator accessor below)
      (child-comparator-accessor-and-below node y-p direction)
      (let ((node-below (funcall below child)))
        (cond ((null node-below) child)
          ((funcall comparator
              (funcall accessor child)
              (funcall accessor node-below))
              child)
          (t node-below)))))

(defun child-comparator-accessor-and-below (node y-p direction)
  "Returns NODE's child in DIRECTION, a node comparison function, a node value accessor, and a below accessor."
  (if (eq direction :right)
      (if y-p
          (values (2d-node-right node)
            #'< #'2d-node-y #'2d-node-lowest-below)
          (values (2d-node-right node)
            #'< #'2d-node-x #'2d-node-lowest-below))
      (if y-p
          (values (2d-node-left node)
            #'> #'2d-node-y #'2d-node-highest-below)
          (values (2d-node-left node)
            #'> #'2d-node-x #'2d-node-highest-below))))

(defun at-or-near-fringe-p (node)
  "Non-null if NODE is on the fringe
   or one joint from the fringe."
  (and (> 2. (2d-node-depth-left node))
       (> 2. (2d-node-depth-right node))))
(defun remove-replacement (replacement parent)
   "Removes REPLACEMENT from its position below NODE, restoring any children of the replacement inside the parent of the replacement. Note that since the replacement is at or near the fringe of the tree, there can never be more than two nodes to reinsert. Since they're reinserted in the node's parent, this operation is always very inexpensive. At worst, we will have to reinsert two nodes with three node comparisons — each reinsertion is compared to the parent and they are compared to each other."
   (let ((progeny (cdr (unwind replacement)))
      ;; the unwinding is minimal, we're at or near fringe.
      (position (position-in-parent replacement parent t)))
      ;; first remove the reference to the replacement in parent.
      (cond ((eq position :left)
        (setf (2d-node-left parent) nil)
        (setf (2d-node-depth-left parent) 0))
        (t (setf (2d-node-right parent) nil)
        (setf (2d-node-depth-right parent) 0)))
      ;; then reinsert the progeny of the replacement.
      (reinsert-nodes parent progeny nil)
      (if (2d-node-balanced-p parent)
        ;; There are too few progeny to require balancing of
        ;; progeny, but the parent may require it.
        parent
        (2d-balance parent))))
;;; Reinsertion code

(defun reinsert-nodes (parent nodes balance-p)
  "Reinserts a list of NODEs.
  (dolist (node nodes)
    (reinsert parent node balance-p)))

(defun reinsert
  (parent node &optional (balance-p t) (balance-direction))
  "Reinserts NODE in PARENT."
  ;;First, forget any tree data in node.
  (strip-2d-node-of-variables node)
  ;;then, simply reinsert it.
  (store-2d-node-1 node parent
   (2d-node-parent parent)
   (2d-node-depth parent)
   nil
   (2d-node-x node)
   (2d-node-y node)
   balance-p
   balance-direction))

;;; Code for ensuring that nodes between the old and new
;;; position of the replacement no longer point to the
;;; replacement as a highest or lowest node below them.

(defun ensure-correct-pointers-below (replacement parent)
  "Walks the tree from REPLACEMENT's former PARENT to the
REPLACEMENT's new position, fixing the below pointers of
any intervening node that points to REPLACEMENT. Since
REPLACEMENT will now appear above these nodes, new below
pointers must be found for these cases. The search for these
new pointers is finitely small, accessing only the children
and grandchildren of each such node."

  ;;get the highest and lowest x and y from the progeny.
  (multiple-value-bind (high-x high-y low-x low-y)
    ;;this form is defined below.
    (generate-below-from-two-generations parent)
  ;;now walk up the tree until you get to the REPLACEMENT,
  ;;replacing any references to REPLACEMENT in each node's
  ;;BELOW slot.
  (ensure-correct-pointers-below-1
   replacement parent high-x high-y low-x low-y
   (oddp (2d-node-depth parent)) nil)))
Internals of ENSURE-CORRECT-POINTERS-BELOW.

```
(defmacro .ensure-best-belows. ()
  "Here we compare the values we already have with other possibilities from the opposite fork. Since we can exploit the BELOW slots of the nodes in the opposite fork, this is not expensive."
  `(when caller
    (multiple-value-bind (high-x high-y low-x low-y)
      (below-candidates-from-opposite-fork caller)
        (when (and high-x
            (or (null highest-x)
              (> (2d-node-x high-x)
                  (2d-node-x highest-x))))
          (setq highest-x high-x))
        (when (and high-y
            (or (null highest-y)
              (> (2d-node-y high-y)
                  (2d-node-y highest-y))))
          (setq highest-y high-y))
        (when (and low-x
            (or (null lowest-x)
              (< (2d-node-x low-x)
                  (2d-node-x lowest-x))))
          (setq lowest-x low-x))
        (when (and low-y
            (or (null lowest-y)
              (< (2d-node-y low-y)
                  (2d-node-y lowest-y))))
          (setq lowest-y low-y))))

(defmacro .check-node-against-belows. ()
  "When recursing in ENSURE-CORRECT-POINTERS-BELOW-1, make sure you carry up the best values."
  `(let ((x (2d-node-x node))
            (y (2d-node-y node)))
      (when (or (null highest-x)
              (> x (2d-node-x highest-x)))
        (setq highest-x node))
      (when (or (null highest-y)
              (> y (2d-node-y highest-y)))
        (setq highest-y node))
      (when (or (null lowest-x)
              (< x (2d-node-x lowest-x)))
        (setq lowest-x node))
      (when (or (null lowest-y)
              (< y (2d-node-y lowest-y)))
        (setq lowest-y node)))
```
;;; The macros on the preceding page expand in this form.

(defun ensure-correct-pointers-below-1
  (replacement node highest-x highest-y
    lowest-x lowest-y y-p caller)
"Does the work for ENSURE-BEST-POINTER-BELOW."
(let ((parent (2d-node-parent node)))
  (adjust-depths node)
  (.ensure-best-belows.)
  (setf (2d-node-lowest-below node)
    (if y-p lowest-x lowest-y))
  (setf (2d-node-highest-below node)
    (if y-p highest-x highest-y))
  (unless (or (null parent) (eq node replacement))
    ;; go no further. This is either the stop node or root.
    (.check-node-against-belows.)
    ;; recurse on NODE's PARENT.
    (ensure-correct-pointers-below-1
      replacement parent highest-x highest-y
      lowest-x lowest-y (not y-p) node))
  ;; a node between the replacement's old and new positions
  ;; may become unbalanced. We do this on the way out of the
  ;; recursion in order to ensure we have the correct values.
  (when (not (2d-node-balanced-p node))
    (adjust-depths-to-root (2d-balance node))))
;;; Code used to extract lowest and highest below whenever
;;; they've been changed because of node repositioning.
;;; Note that these procedures look only so far as a node's
;;; grandchildren. Also note that when walking UP the tree
;;; from the old to the new positions of the replacement,
;;; we carry back the highest and lowest Xs and Ys seen so
;;; far, so that search to the grandchildren of subsequent
;;; nodes need only examine the two grandchildren of the
;;; other fork. ;;;

(defun generate-belows-from-two-generations (node)
  "Returns the highest x and y and the lowest x and y below NODE.
  It does so by comparing NODE's children, grandchildren, and their
  lowers."
  (multiple-value-bind (l-high-x l-high-y l-low-x l-low-y)
      (below-candidates-of-child (2d-node-left node))
    (multiple-value-bind (r-high-x r-high-y r-low-x r-low-y)
        (below-candidates-of-child (2d-node-right node))
      ;; cons on the stack, create zero garbage.
      (stack-let ((high-x '(),1-high-x ,r-high-x))
        (high-y '(),1-high-y ,r-high-y))
      (low-x '(),1-low-x ,r-low-x)
      (low-y '(),1-low-y ,r-low-y))
    (values (2d-node-compare #'> '#2d-node-x high-x)
     (2d-node-compare #'> '#2d-node-y high-y)
     (2d-node-compare #'< '#2d-node-x low-x)
     (2d-node-compare #'< '#2d-node-y low-y))))

(defun below-candidates-from-opposite-fork (caller)
  "Returns the highest x and y and the lowest x and y
  in the opposite fork of the parent of the CALLER."
  (multiple-value-bind (high-x high-y low-x low-y)
      (below-candidates-of-child (sibling caller))
    (values high-x high-y low-x low-y))
(defun below-candidates-of-child (child)
  (when child
    (let* ((y-p (oddp (2d-node-depth child)))
            (accessor (if y-p #'2d-node-y #'2d-node-x))
            (my-accessor (if y-p #'2d-node-x #'2d-node-y))
            (multiple-value-bind (highest lowest)
                (values-from-grandchildren child accessor)
                ;; cons on the stack.
                (stack-let ((high `((child
                                      ,(2d-node-highest-below child)))
                             (low `((child
                                     ,(2d-node-lowest-below child))))
                (let ((chighest
                       (2d-node-compare #'> my-accessor high))
                       (clowest
                       (2d-node-compare #'< my-accessor low)))
                (if y-p
                    (values chighest highest clowest lowest)
                    (values highest chighest lowest clowest))))))))

(defun values-from-grandchildren (child accessor)
  "Returns the nodes with the highest and lowest values in the tree with CHILD as root, using ACCESSOR to control the dimension of comparison."
  (stack-let*
    (left-grandchild (2d-node-left child))
    (right-grandchild (2d-node-right child))
    (left-highest-below
      (if left-grandchild
        (2d-node-highest-below below left-grandchild)))
    (right-highest-below
      (if right-grandchild
        (2d-node-highest-below right-grandchild)))
    (left-lowest-below
      (if left-grandchild
        (2d-node-lowest-below below left-grandchild)))
    (right-lowest-below
      (if right-grandchild
        (2d-node-lowest-below right-grandchild)))
    (high-candidates
      `((child ,left-grandchild ,right-grandchild
        ,left-highest-below ,right-highest-below))
    (low-candidates
      `((child ,left-grandchild ,right-grandchild
        ,left-lowest-below ,right-lowest-below))
    (values (2d-node-compare #'> accessor high-candidates)
             (2d-node-compare #'< accessor low-candidates))))
;; Splicing out a node from the tree.

(defun strip-2d-node-of-variables (node)
  (setf (2d-node-parent node) nil)
  (setf (2d-node-left node) nil)
  (setf (2d-node-right node) nil)
  (setf (2d-node-depth node) 0.)
  (setf (2d-node-depth-right node) 0.)
  (setf (2d-node-depth-left node) 0.)
  (setf (2d-node-lowest-below node) nil)
  (setf (2d-node-highest-below node) nil)
  node)

;; Splicing in a replacement.

(defun swap-node-for-replacement (node replacement)
  "Fills REPLACEMENT with NODE's data."
  (let* ((left (2d-node-left node))
         (right (2d-node-right node))
         (parent (2d-node-parent node))
         (rparent (2d-node-parent replacement))
         (rposition
          (if rparent
              (position-in-parent replacement rparent))))
    (when rposition
      (if (eq rposition :left)
          (setf (2d-node-left parent) nil)
          (setf (2d-node-right parent) nil)))
    (setf (2d-node-left replacement) left)
    (setf (2d-node-right replacement) right)
    (when left (setf (2d-node-parent left) replacement))
    (when right (setf (2d-node-parent right) replacement))
    (when parent
      (if (eq :left (position-in-parent node parent t))
          (setf (2d-node-left parent) replacement)
          (setf (2d-node-right parent) replacement)))
    (setf (2d-node-parent replacement) parent)
    (setf (2d-node-depth replacement) (2d-node-depth node))
    (setf (2d-node-depth-left replacement)
          (2d-node-depth-left node))
    (setf (2d-node-depth-right replacement)
          (2d-node-depth-right node)))
Node utilities.

(defun create-2d-node (x y)
  "Returns a 2D-NODE with X and Y values."
  (let ((node (make-2d-node)))
    (setf (2d-node-x node) x)
    (setf (2d-node-y node) y)
    node))

(defun 2d-node-attach (node parent direction)
  "Attaches NODE to PARENT in DIRECTION."
  (setf (2d-node-parent node) parent)
  (cond ((eq direction :left)
         (setf (2d-node-left parent) node))
        ((eq direction :right)
         (setf (2d-node-right parent) node))
        (t (error "Direction is " direction))))

(defun maximum-depth (node)
  "Returns the maximum depth below NODE."
  (max (2d-node-depth-left node)
       (2d-node-depth-right node)))

(defun position-in-parent
  (child &optional (parent (2d-node-parent child)) error-p)
  "Returns CHILD's position in PARENT, either :RIGHT or :LEFT. If ERROR-P is non-null, an error will be signalled if CHILD is not a child of PARENT. If ERROR-P is null, NIL will be returned if CHILD is not PARENT's child."
  (cond ((eq child (2d-node-left parent)) :left)
        ((eq child (2d-node-right parent)) :right)
        (error-p (error "S is not a child of " child parent))))
;;; More node utilities

(defun sibling (node &optional (parent (2d-node-parent node)))
  "Returns NODE's sibling, or NIL if there isn't one.
Also returns the sibling's parent as the second value."
  (when parent
    (let ((left (2d-node-left parent)))
      (when left
        (if (eq left node)
            (values (2d-node-right parent) parent)
            (values left parent))))))

(defun unwind (node)
  "Returns a list of NODE and all its progeny."
  (cond ((null node) nil)
        (t (cons node
            (nconc (unwind (2d-node-left node))
                (unwind (2d-node-right node)))))))

(defun 2d-node-root (node)
  "Returns the root of NODE's tree."
  (let ((parent (2d-node-parent node)))
    (cond ((null parent) node)
          (t (2d-node-root parent))))))

(defun 2d-node-compare (predicate accessor node-list)
  "Compares pairwise the nodes in NODE-LIST, returning the NODE whose slot, accessed by ACCESSOR, best satisfies the PREDICATE."
  (let ((result (car node-list)))
    (dolist (x (cdr node-list) result)
      (and x
        (or (null result)
          (funcall predicate
           (funcall accessor x)
           (funcall accessor result)))
        (setq result x))))))
;; Still more node utilities.

(defun adjust-depths (node)
  "Adjusts the depths below of NODE."
  (let* ((left (2d-node-left node))
         (right (2d-node-right node)))
    (setf (2d-node-depth-left node)
          (or (and left
                   (+ 1. (maximum-depth left)))
              0.))
    (setf (2d-node-depth-right node)
          (or (and right
                   (+ 1. (maximum-depth right)))
              0.))))

(defun adjust-depths-to-root (node)
  (cond ((null node) nil)
        (t (adjust-depths node)
           (adjust-depths-to-root
                    (2d-node-parent node)))))
;;; Range searching.

;;; Using balanced 2D trees, all the elements in any region in the two-dimensional space are extracted in time proportional to $E + \log N$, where $E$ is the number of elements found and $N$ is the number of nodes in the tree.

;;; The top level form is 2D-RANGE-SEARCH.

(defun 2d-range-search (x low x nil)
  (let ((high-x* nil)
        (low-x* nil)
        (high-y* nil)
        (low-y* nil)

        (defmacro with-range-search-bounds
          (low-x high-x low-y high-y &body body)
          "Dynamically binds some special vars to avoid passing around arguments."
          (let (((*high-x* ,high-x)
                  (*low-x* ,low-x)
                  (*high-y* ,high-y)
                  (*low-y* ,low-y))
                ,body))

        (defun range-relation (node)
          (values (relation 2d-node-x node) *low-x* *high-x*)
          (relation 2d-node-y node) *low-y* *high-y*))

        (defun relation (value low high)
          "Returns a keyword indicating whether VALUE is :ABOVE :WITHIN or :BELOW the range denoted by LOW and HIGH."
          (cond ((and low (> low value) :below))
                ((and high (> value high) :above))
                (t :within)))

)}
;;; Procedures for searching temporal relations.

(defun 2d-node-x-y (node)
  (values (2d-node-x node) (2d-node-y node)))

(defun time-superiors (node &key (tree *root*)
  accept-same-starts
  accept-same-ends)
  "Find the time-superiors of NODE."
  (multiple-value-bind (x y) (2d-node-x-y node)
    (unless accept-same-starts (decf x))
    (unless accept-same-ends (incf y))
    (2d-range-search :low-x 0 :high-x x
                     :low-y y :tree tree)))

(defun time-inferiors (node &key (tree *root*)
  accept-same-starts
  accept-same-ends)
  "Find the time-inferiors of NODE."
  (multiple-value-bind (x y) (2d-node-x-y node)
    (unless accept-same-starts (incf x))
    (unless accept-same-ends (decf y))
    (2d-range-search :low-x x :high-x y
                     :low-y x :high-y y
                     :tree tree)))
(defun time-overlaps (node &key (tree *root*))
   (accept-same-starts
    (accept-same-ends
     (accept-starts-at-ends))
   "Find the time-overlaps of NODE."
   (if (and accept-same-starts accept-same-ends)
     (error "S cannot be an overlap of itself." node))
   (multiple-value-bind (x y) (2d-node-x-y node)
     (let ((high-x (unless accept-starts-at-ends (- y 1.)))
       (unless accept-same-starts (incf x))
       (unless accept-same-ends (incf y))
       (2d-range-search :low-x x :high-x high-x
                        :low-y y :tree tree)))

(defun time-underlaps (node &key (tree *root*))
   (accept-same-starts
    (accept-same-ends
     accept-ends-at-starts)
   "Find the time-underlaps of NODE."
   (if (and accept-same-starts accept-same-ends)
     (error "S cannot be an underlap of itself." node))
   (multiple-value-bind (x y) (2d-node-x-y node)
     (let ((low-y (unless accept-ends-at-starts (+ x 1.)))
       (unless accept-same-starts (decf x))
       (unless accept-same-ends (decf y))
       (2d-range-search :low-x 0 :high-x x
                        :low-y low-y :high-y y
                        :tree tree))))
(defun time-disjoint-before (node &key (tree *root*))
  "Find the time-disjoint-before NODE."
  (let ((x (2d-node-x-y node)))
    (2d-range-search :low-x 0 :high-x (- x 1)
                     :low-y 0 :high-y (- x 1)
                     :tree tree)))

(defun time-disjoint-after (node &key (tree *root*))
  "Find the time-disjoint-after NODE."
  (let ((y (2d-node-y node)))
    (2d-range-search :low-x (+ 1 y) :low-y (+ 1 y)
                     :tree tree)))

(defun time-same-starts (node &key (tree *root*))
  "Find the time-same-starts as NODE."
  (let ((x (2d-node-x node)))
    (2d-range-search :low-x x :high-x x :low-y x
                     :tree tree)))

(defun time-same-ends (node &key (tree *root*))
  "Find the time-same-ends as NODE."
  (let ((y (2d-node-y node)))
    (2d-range-search :low-x 0 :high-x y :low-y y
                     :high-y y :tree tree)))

(defun time-ends-at-starts (node &key (tree *root*))
  "Find the time-ends-at-starts of NODE."
  (let ((x (2d-node-x node)))
    (2d-range-search :low-x 0 :high-x x
                     :low-y x :high-y x
                     :tree tree)))

(defun time-starts-at-ends (node &key (tree *root*))
  "Find the time-starts-at-ends of NODE."
  (let ((y (2d-node-y node)))
    (2d-range-search :low-x y :high-x y
                     :low-y y :tree tree)))

(defun time-equality (node &key (tree *root*))
  "Find the nodes equivalent to NODE."
  (multiple-value-bind (x y) (2d-node-x-y node)
    (2d-range-search :low-x x :high-x x
                     :low-y y :high-y y :tree tree)))