MASSACHUSETTS INSTITUTE OF TECHNOLOGY
ARTIFICIAL INTELLIGENCE LABORATORY

A.I. Memo No. 871 May 1986 Revised

Hermeneutics:

From Textual Explication to Computer Understanding?

John C. Mallery, Roger Hurwitz, Gavan Duffy

Published in *The Encyclopedia of Artificial Intelligence*, Stuart. C. Shapiro, editor, John Wiley & Sons, New York, 1987.

Abstract

Discussions of the theoretical foundations of artificial intelligence increasingly refer to hermeneutics, a branch of continental European philosophy concerned with human understanding and the interpretation of written texts. Dreyfus and Winograd draw heavily on hermeneutics to question the feasibility of AI and cognitive science. But, hermeneutics also offers insights that may contribute to the understanding of meaning, translation, architectures for natural language understanding, and even to the methods suitable for scientific inquiry in AI. After briefly reviewing the historical development of hermeneutics as a method of interpretation, from its classical use through the modern debates, this article examines the contributions of hermeneutics to the human sciences. This background provides perspective for a review of recent hermeneutically-oriented AI research. This includes the Alker, Lehnert and Schneider computer-assisted techniques for coding the affective structures of narratives, the earlier positive proposal by Winograd, and subsequently Bateman, the later pessimistic Winograd and Flores on the possibility of AI, as well as the system-building efforts of Duffy and Mallery.

Keywords: Bootstrapping, Cognitive Science, Dialectics, Discourse, Eidetic, Epistemology, Explanation, Hermeneutic Circle, Hermeneutics, Interpretation, Metaphor, Ontology, Phenomenology, Philosophical Foundations, Philosophy of Science, Text, Understanding.

Contents

1 Introduction

Recent debates about the theoretical foundations of Artifical Intelligence (AI) refer to hermeneutics, the branch of continental philosophy which treats the understanding and interpretation of texts. Applying certain hermeneutic insights, Dreyfus [1], Winograd [2] and Winograd and Flores [3] have questioned the functionalist cognitive science paradigm that guides most contemporary AI research, particularly in natural language processing and common-sense reasoning. Dreyfus draws upon the hermeneutic philosophy of Martin Heidegger [4] to deny the possibility of formalizing mental processes, and therefore, creating artificial intelligences.[1] Winograd and Flores reach a similar conclusion based on a hermeneutically-informed technical argument. Yet, in addition to being a source of doubts, hermeneutics may illuminate problems like the nature of meaning and understanding and thereby help reconstruct the functionalist paradigm [2].

To help clarify the relevance of hermeneutics for AI research, this article first reviews the major strains of hermeneutic thought. These positions include the naive hermeneutics of early modern Europe and Dilthey's more historically conscious, 19th century *methodological hermeneutics*, which sought to produce systematic and scientific interpretations by situating a text in the context of its production. In the twentieth century, Heidegger's and Gadamer's *philosophical hermeneutics* shifted the focus from interpretation to existential understanding, which was treated more as a direct, non-mediated, authentic way of being in the world than as a way of knowing. Reacting to the relativism of this position, Apel and Habermas introduced *critical hermeneutics* – a methodologically self-reflective and comprehensive reconstruction of the social foundations of discourse and inter-subjective understanding. Finally, Ricoeur in his *phenomenological hermeneutics* attempted to synthesize the various hermeneutic currents with structuralism and phenomenology.

This background situates AI researchers and critics who draw from the various hermeneutic traditions. In their investigations of the affective structure of texts and in their concern with systematic rules for identifying the genres of narrative, Alker, Lehnert, and Schneider [5] in effect pursue a classical hermeneutical program tempered by phenomenological hermeneutics. Other researchers [2, 6, 7] draw from philosophical hermeneutics to propose strategies for developing computer systems that understand natural language. A third approach [3], aligned with philosophical hermeneutics, argues that computer understanding of natural language is exceedingly difficult and probably intractable. A fourth group [8] has developed an implementation guided in part by ideas from phenomenological hermeneutics but informed by the other variants as well.

Hermeneutic theories differ in several characteristic ways from approaches to meaning and understanding which are better known to AI researchers. Hermeneutics grounds the meaning of texts in the intentions and histories of their authors and/or in their relevance for readers. In contrast, analytic philosophy usually identifies meaning with the external referents of texts and structuralism finds meaning in the arrangement of their words. Hermeneutics regards texts as means for transmitting experience, beliefs and judgments from one subject or community to another. Hence the determination of specific meanings is a matter for practical judgement and common sense reasoning – not

[1] In a personal communication (March, 1986), Dreyfus indicated that he has recently moderated his views and now considers AI very difficult but not necessarily impossible.

for *a priori* theory and scientific proof. This attitude reflects the origin of hermeneutics in ancient-world efforts to determine systematically the meaning, intent and applicability of sacred and legal texts. Hermeneutic theories and applications also share the idea of the *hermeneutic circle* or the notion that understanding or definition of something employs attributes which already presuppose an understanding or a definition of that thing. Circles or spirals of understanding arise in interpreting one's own language, a foreign language or an observed action, in confirming a theory and in distinguishing between background knowledge and facts [9]. The existence of these circularities raises questions for hermeneutics regarding the grounding and validity of understanding.

The philosophical concept of the hermeneutic circle resembles the distinctly computational notion of bootstrapping - a process which uses a lower order component (a *bootstrap component*) to build a higher order component that is used in turn to reconstruct and replace the lower order component. Bootstrapping has been introduced in the design of certain knowledge bases [10, 11, 12] and in AI-oriented theories of cognitive development [13, 14, 15, 16], and should be distinguished from hierarchical layering in systems which do not include the "strange loop" of replacing the bootstrap component. The similarity of the hermeneutic circle and bootstrapping suggests the possibility of an important contribution from hermeneutics to AI architectures for natural language processing and for common-sense reasoning.

2 Classical Methodological Hermeneutics

2.1 Origins

Hermeneutics as a general science of interpretation can be traced back to more domain-specific applications in the ancient Greeks' study of literature and in ancient Biblical exegesis.[2] Regarding texts as organic or coherent wholes, rather than collections of disjointed parts, the Greeks expected a text to be consistent in grammar, style and ideas. Accordingly, they codified rules of grammar and style that they used to verify and emend textual passages. By extending the logic of part and whole to a writer's or school's entire output, the Greeks were also able to attribute works with uncertain origin. Although the Jewish Rabbis and the early Church Fathers deployed similar philological tools, their Biblical exegeses were better known for the development of allegorical readings, frequently at the expense of the texts' literal meaning. Their interpretations found within the visible sign a hidden sense in agreement with the intention which they beforehand ascribed to the text. Since instances of this method are found for the Vedas, Homer, the Koran and other sacred writings, it seems a typical strategy for reconciling an enlightened or moral world-view with texts whose "outward" earthiness or banality seems beneath the dignity of the gods being celebrated [18].

The Middle Ages witnessed the proliferation of non-literal interpretations of the

[2]The word *hermeneutics* was coined in the 17th century [17] on the basis of the Greek *hermeneuein*, "to interpret," which signified equally a declamation of a text, an explanation of a situation or a translation from a foreign tongue. *Hermeneuein* itself derived from the name of Hermes, the winged messenger god of ancient Greece, who both delivered and explained the messages of the other gods.

Bible. Christian commentators could read Old Testament stories simultaneously as precursors of analogous episodes in the New Testament, symbolic lessons about Church institutions, and allegories about spiritual traits [19]. In each case, the meaning of the signs was constrained by imputing a particular intention to the Bible, such as teaching morality, but these interpretive bases were posited by the religious tradition rather than suggested by a preliminary reading of the text. Thus, when Martin Luther argued that Christians could rediscover their faith by reading the Bible themselves, Catholic Church officials not surprisingly responded that the Bible was too obscure to read without their guidance. The Protestant exegesis, which appeared after Luther's translation of the Bible, tended to view the texts as responses to historical or social situations rather than expressions of theological principles. Assuming that the New Testament documented the Christian faith, one reader's guide proposed that contradictory statements and difficult passages in the New Testament could be clarified by comparing their possible meanings with contemporaneous Christian practices. The example suggests that interpretation might rely on *empathetic understanding*, the interpreter's self projection into the author's space. Indeed, it was just such empathy that Schleiermacher and Dilthey raised to a methodological principle in their attempt to create a general hermeneutics.

2.2 The Methodological Hermeneutics of Schleiermacher and Dilthey

Schleiermacher [20] proposed to join classical philology's focus on grammar and style and Biblical exegesis' concern for themes, creating a general hermeneutics with principles independent of domain specific interpretation principles. Schleiermacher compared the reader's approach to a text with the efforts by participants in a dialogue to understand each other, and he depicted the dialogue in terms of a speaker who puts together words to express his thoughts and a listener who understands this speech as part of a shared language and as part of the speaker's thinking [21: 37]. The listener can comprehend the words and sentences because they are drawn from the language's lexicon and follow its grammatical rules, but the listener can also recognize the intentions behind the words by virtue of being in the same situation and sharing a common human nature with the speaker. Since Schleiermacher's concept of understanding includes empathy (projective introspection) as well as intuitive linguistic analysis, it is much richer than the idea in modern communication theories that understanding is merely the decoding of encoded information. Interpretation is built upon understanding and has a grammatical, as well as a psychological moment. The grammatical thrust has a bootstrapping flavor: It places the text (or expression) within a particular literature (or language) and reciprocally uses the text to redefine the character of that literature. The psychological thrust is more naive and linear. In it, the interpreter reconstructs and explicates the subject's motives and implicit assumptions. Thus Schleiermacher claimed that a successful interpreter could understand the author as well, as or even better than, the author understood himself because the interpretation highlights hidden motives and strategies.

Broadening Schleiermacher's hermeneutics, Dilthey [22] developed a philosophy of method for history and the human sciences that he believed could produce objective knowledge but avoid the reductionist, mechanistic, ahistorical explanatory schema of the natural sciences. Dilthey argued that texts, verbal utterances, art and actions were meaningful expressions whose "mental contents" or intentions needed to be comprehended. He claimed that investigating human interactions was more like interpreting a poem or discourse, than doing physics or chemistry experiments [23]. Dilthey termed the desired comprehension of events and expressions "understanding" (*verstehen*) and attempted to distinguish it from the explanatory knowledge (*erkennen*) generated by the hypothetico-deductive method of the natural sciences.

Dilthey initially followed Schleiermacher in identifying understanding as empathy guaranteed by the notion of a common human nature. Although he recognized that the outlook and values of people varied over different historical periods and cultures, Dilthey argued that, because historians themselves thought and acted, they could relive and understand what people in the past were trying to express and accomplish in their writings, speeches, actions and art. Nevertheless, many of his contemporaries criticized this position because it relied on introspection and an underspecified, non-critical psychology. Stung by this criticism and influenced by the neo-Kantian idea that works of art and literature embodied the formal values of their respective periods, Dilthey revised his position. He began to emphasize that texts and actions were as much products of their times as expressions of individuals, and their meanings were consequently constrained by both an orientation to values of their period and a place in the web of their authors' plans and experiences. In this revision, meanings are delineated by the author's *weltanschauung*, or world-view reflecting a historical period and social context. *Understanding* (*verstehen*), the basis for methodological hermeneutics, involves tracing a circle from text to the author's biography and immediate historical circumstances and back again. *Interpretation*, or the systematic application of understanding to the text, reconstructs the world in which the text was produced and places the text in that world.[3]

This circular process precludes an interpretation of a text from being unique and scientifically objective, like the explanation of a chemical reaction, inasmuch as knowledge of the author's or agent's world may itself critically depend on the present interpretation [9]. Dilthey and his recent followers, Hirsch [26] and Betti [27], claim, however, that interpretations become more valid as they assimilate more knowledge about the author and the author's values, instead of reflecting the interpreter's own values or sense of reality. Dilthey's method in effect bootstraps from a whole (a biography, a set of works) whose themes may be repeatedly respecified through the elaboration of one of its parts (the action or work). The process eventually reaches stability because successive interpretations of the work or action serve to constrain subsequent refinements in the background model of the author. The strength and validity of such constraints depends on the currency and robustness of that model. Increases in temporal and cultural distance between the speaker and interpreter decrease the reliability of interpretation, but this neither forecloses the possibility of such a model nor denies the potential for a valid interpretation.

[3]See Dilthey [23] for a sampling of Dilthey's writings on history and the human sciences, and Ermarth [24] and Plantinga [25] for their discussion.

3 Philosophical Hermeneutics

3.1 Heidegger's Ontological Hermeneutics

In *Being and Time* [4], Heidegger undermines the notion of objectivity in Husserl's phenomenology[4] [31], and by extension, in methodological hermeneutics. Husserl argues that objective interpretation is possible using his transcendental phenomenological method that requires bracketing the subjectivity inhering in the interpreter's *life-world* (*Lebenswelt*), the world of personal experience and desires. Heidegger denies that this bracketing is possible. He claims instead that the understanding of a situation is directly mediated by a fore-knowledge, or sensitivity to situations, that is comprised by the understander's life-world. Therefore, suspending that life-world would preclude the possibility of understanding, altogether. Heidegger reaches his conclusion by contending that as a necessary part of human "being-in-the-world" (*Dasein*), things are perceived according to how they are encountered and used in one's everyday routines and tasks. Perception and apprehension thus move from fore-knowledge to an *existential understanding*, a largely unreflective and automatic grasp of a situation that triggers a response. This understanding must be incomplete because *Dasein* is both historical and finite. It is historical in that understanding builds from the fore-knowledge accumulated from experience. It is finite due to "throwness," the necessity of acting in situations without the time or ability to grasp the full consequences of actions or plans in advance. Only when actions fail to meet the exigencies of the situation and "breakdown" occurs, do individuals stand back and assume the theoretical attitude of science which sees things "objectively," as discrete objects separate from the self and resistant to one's will.

Heidegger brings hermeneutics from a theory of interpretation to a theory of existential understanding. He "depsychologizes" hermeneutics by dissociating it from the empathetic perception of other beings. Understanding now appears as a *no-longer-conscious* component of *Dasein*; it is embedded within the context of specific situations and plans, with, in effect, finite computational resources. Therefore, interpretation (*Auslegung*) which depends on such existential understanding (*Verstehen*) is not the general logical method found in classical philology, but refers to a conscious recognition of one's own world. Dilthey's methodological hermeneutic circle is consequently supplanted by the more fundamental *ontological hermeneutic circle*, which leads from existential understanding situated in a world to a self-conscious interpretive stance. This self-consciousness, however, cannot escape its limitations to achieve a transcendental understanding in the sense of Hegel [32, 33], who considered rationality the ability to reflectively accept or reject (transcend) the received socio-cultural tradition [33.5]. According to this reading of Heidegger, fore-knowledge is accumulated over time and constrains successive exercises of existential understanding. But self-conscious understanding cannot choose which elements in the experience based foreknowledge are respecified in the bootstrapping process.[5]

[4] For readers unfamiliar with phenomenology, Schmitt [28] and Zaner [29] present concise overviews and Ricoeur [30] provides an extensive analysis.

[5] Green [34] presents a concise overview of Heidegger's contributions to philosophy. Steiner [35] and Palmer [17: 124-161] provide accessible introductions to Heidegger's thought. Murray [36] contains an informative collection of essays discussing Heidegger's thought.

3.2 Gadamer's Philosophical Hermeneutic

In his philosophical hermeneutics, Gadamer [37] follows his teacher Heidegger in recognizing that the ties to one's present *horizons*, one's knowledge and experience, are the productive grounds of understanding. However, Gadamer argues that these limits can be transcended through exposure to others' discourse and linguistically encoded cultural traditions because their horizons convey views and values that place one's own horizons in relief.[6] He stresses the role of language in opening the subject to these other subjectivities and their horizons. In forcefully stressing the role of language in opening the subject to other subjectivities in constituting traditions, Gadamer places language at the core of understanding.[7] Consequently, understanding for Gadamer does not scientifically reconstruct a speaker's intention, but instead mediates between the interpreter's immediate horizon and his emerging one.

For Gadamer, understanding is bound and embedded in history because understanding deploys the knower's *effective-history*, personal experience and cultural traditions, to assimilate new experiences. Thus, the initial structure of an effective-history constrains the range of possible interpretations, excluding some possibilities and calling forth others. As effective-history constitutes the *prejudices* brought to bear in understanding, it simultaneously and dialectically limits any self-conscious attempts to dissolve those prejudices. Gadamer thus explicitly opposes the scientific ideal of prejudiceless objectivity in interpretation. In this respect, he moves beyond Heidegger, who regarded so-called scientific objectivity as a derivative of existential understanding. Gadamer does not deny the importance of either *scientific understanding* or *critical interpretation*, a form of interpretation that introspectively questions assumptions unreflectively inherited from cultural traditions. His focus on the human context of knowledge emphasizes the need for repeated attempts at critical understanding, through which people can gain the insight needed to correct their prejudices. But, if prejudices may be individually overcome, their fact is inescapable. It imposes *a priori* limitations on the extent to which a self-reflective methodology can eliminate distortions from scientific inquiry. The *critical self-consciousness* of a rational agent who introspectively questions received traditions may counter distorting consequences of effective-history, but it at best only leads to successive approximations of objectivity.[8]

The resulting theory of meaning differs from the methodological hermeneutics of Schleiermacher and Dilthey, which identifies the meaning of a text with its author's intentions and which seeks to decipher the text by uncovering the world-view *behind* it. For Gadamer, understanding recreates the initial intention embodied in the text, by elucidating the subject matter that the text addresses (its *aboutness*). The process

[6] Green [34] contends that Heidegger fails to show how the historicity of the individual relates to the history of a broader community.

[7] Gadamer's [37.4] position approximates the hypothesis advanced by the American linguists Sapir [37.5] and Whorf [37.6] which holds that the individual's language partially determines his or her conceptual system and world view. According to the hypothesis, complete translations between languages is impossible, and understanding another language requires complete immersion, accompanied by a change in thinking.

[8] Gadamer's position prompts the philologists Betti [27] and Hirsch [26] to complain that its relativism destroys all bases for validating an interpretation and so defeats the purpose of interpretation. Social theorist Habermas [38] also criticizes Gadamer's relativism. See below.

moves the text beyond its original psychological and historical contexts and gives it a certain "ideality" of meaning, which is elaborated in a dialogue between the interpreter and the text. The dialogue is grounded in the concern which the interpreter and the author share toward a common question and a common subject matter. In confronting a viewpoint reflecting a different set of horizons, the interpreter can find his own horizons highlighted and reach critical self-consciousness. In seeking the key question, the interpreter repeatedly transcends his own horizons while pulling the text beyond its original horizons until a fusion of the two horizons occurs. The interpreter's imagination can also play a role in the dialogue with texts and carry the understanding of the subject matter beyond the finite interpretation realized in methodological hermeneutics. Nevertheless, the interpretations are constrained by the questions posed, since each question calls forth frameworks within which the subject matter must be understood. The meaning of a text then is not fixed, but changes over time according to how it is received and read. Thus for Gadamer, to understand is to understand differently than the author or even one's own earlier interpretations, precisely because the process involves creating new horizons by bootstrapping from the old horizons which they replace. But the notion of bootstrapping in Gadamer moves beyond the one in Heidegger because Gadamer allows prejudices to come into a conscious focus that may direct their individual supersession.

Gadamer does not merely work through Heidegger's philosophical program. He also redirects philosophical hermeneutics along partly Hegelian lines by appropriating substantial parts of the Hegelian transcendental philosophy that Heidegger eschewed. Gadamer's concepts of the openness of language and the ability of people to transcend their interpretive horizons are based on Hegel's dialectic of the limit, in which the recognition of limits constitutes the first step in transcending them. The concept of understanding as a concrete fusing of horizons is derived ultimately from Hegel's idea that every new achievement of knowledge is a mediation, or a refocusing of the past within a new, present situation [39: xxxix-xl], which attempts to explain mind and logic on the basis of the dialectical resolution of more basic and antithetical concepts [32]. As each opposition is resolved, the resulting synthesis is found to be opposed to yet another concept and that opposition must also be dialectically resolved. This purely subjective and continual unfolding interacts with and is conditioned by experience, particularly the experience of language, which tends to mold the developing subject in conformity with the traditions encoded in linguistic utterances and in the language itself. However, Gadamer clearly resists Hegel's notion of the self-objectifying, transcendental subject. Instead, he views the logical and ontological categories with which Hegel marks the unfolding of thought as distillations of the logic inherent in language, particularly the German language, whose use as a tool for speculative philosophy Hegel brought to perfection [40]. This view affirms the relativist position that thought and reason are always determined by the historical traditions of a linguistic community [41].

4 Critical Hermeneutics

4.1 Strategic Orientation

Heidegger's and Gadamer's critique of objectivity was particularly challenging for social theorists because empirical social science and normative social theory depend ultimately on the characterization of events and situations. At a minimum, the practical need to assess truth-claims and interpretations had to be reconciled with the critique of objectivity. Apel [42] and Habermas [43, 44] sought the means for the reconciliation in conjoining methodological hermeneutics with ordinary language philosophy. Their point of departure was the "critique of ideology," originated by Marx, which argues that beliefs and ideas reflect the holders' social class interests. Although implying that an objective social reality might ultimately be described view also helps explain conflicts in beliefs among members of the same society. Armed with it, Apel and Habermas could conceive of a hermeneutically inspired analysis of communication in linguistic communities. Thus, just as Heidegger's ontological hermeneutic concentrates on the individual's apperception of experience, from the inside out, critical hermeneutics concentrates on individuals situated in groups, from the outside in.

Apel and Habermas argue that of the three divisions of the study of language – syntactics, semantics, and pragmatics – only the first two have been adequately studied by the school of ordinary language philosophy descending from Wittgenstein [45]. They believe that no account of human understanding can be believed if explained as a theory about a single, asocial and ahistorical being. On the contrary, understanding may only be explained by reference to the social and historical setting in which understanding occurs and in the discursive or dialogical situation in which communication takes place. Truth and meaning do not await discovery, but are negotiated by actors who come to consensus on issues of truth and meaning through social discourse. This perspective may be contrasted with the first principles of research programs, such as Chomsky's [46, 47, 48], which seek to explicate language use and language learning on the basis of an examination of a monological model of the competence of an ideal speaker-hearer abstracted from its social situation [49]. Although studies of syntax and semantics are surely necessary for an adequate grasp of the human linguistic faculty, they are by no means sufficient. Any adequate understanding of language, Habermas [50, 51] asserts, must be grounded in the practical purposes for which speakers use language.

4.2 Universal Pragmatics

To provide such grounding, Habermas [50] proposed a universal pragmatics,[9] the primary task of which is the identification and reconstruction of the necessary preconditions for the possibility of understanding in discursive communication. Turning to ordinary language philosophy, he attempts this reconstruction by linking Austin's [53] and Grice's [54] notions of felicity conditions underlying discourse to Searle's [55] theory of speech acts and to a consensus theory of truth, which holds that truth claims are resolved through

[9] Thompson [52] provides a short overview and discussion of universal pragmatics.

reasoned discussion culminating in consensus. Habermas does not confine universal pragmatics to the analyses of language and speech. Rather, because he sees language as the medium in which all human action is explicated and justified, he intends "universal pragmatics" as the groundwork for a general theory of social action.

The resulting critical hermeneutics holds that intersubjective communication is possible, despite differences in the participants' pre-understandings, because the participants in effect posit as an ideal the attainment of a consensus (concerning the validity of statements) that is free from constraints imposed upon them by others and from constraints that they might impose upon themselves. That is, a participant posits a situation in which all participants can freely try to convince others (or be convinced by them) and in which all have equal opportunity to take dialogue roles. Put another way, participation in dialogue admits the possibility of reinterpreting and changing the perceived situation. Habermas and Apel term this idealization the *ideal speech situation* and consider it the participants' emancipatory interest – the situation of freedom to which they aspire. This ideal might never be attained, but even to approach it, the participants must overcome *systematically distorted communication*, which suppresses and conceals the speakers' interests. According to Habermas, these distortions are produced by the division of labor and a correlated structure of domination. Habermas turns to a Freudian psychotherapeutic model to prescribe treatment for the pathological consequences of the systematically distorted horizons produced under these conditions. According to him, the task of the social theorist is to act as therapist, encouraging citizens (patients) to reject internalizations of distorted institutional arrangements (class domination). For Habermas, then, understanding involves compensating for these distortions, and interpretation requires an account of how they were generated.

4.3 The Habermas-Gadamer Debate

Gadamer [56] attacks this position by pointing out that the psychotherapist or social theorist is not immune from the pre-understandings of tradition and that these pre-understandings are not themselves necessarily free of distortion. Gadamer sees Habermas' effort as part of the traditional social scientific goal of attaining "objective" knowledge of the social realm. Habermas [38] appears to believe that the social theorist, like Schleiermacher's interpreter, can understand the social actor better than the social actor understands himself. That is beyond belief for Gadamer, given his notion of ontological pre-understanding. For his part, Habermas sees Gadamer as too ready to submit to the authority of tradition and too reticent to offer any methodological considerations (apart from the exceedingly abstract notion of "interpretive horizons"), thereby giving unwitting support to positivist degradations of hermeneutics.

In reply to Gadamer's claim that prejudices are inescapable, Habermas insists that a *self-reflective* methodology can overcome prejudices and that an objective social theory can be approached, bootstrapping from an initial understanding of society. Habermas argues that the systematic distortions in communication which bias an initial understanding of society can be analyzed and reduced using generalization from empirical knowledge of society, quasi-causal explanation (deductive verification), and historical critique. To build this comprehensive social theory, Habermas must provide a theory of knowledge grounded in: 1) a general theory of communicative action; 2) a general theory of socialization to explain the acquisition of the competence that underpins communicative action; 3) a theory of social systems to show the material constraints on socialization and their reflection in cultural traditions; 4) a theory of social evolution that allows theoretical reconstruction of the historical situations in which communicative action obtains. But this move apparently fails to counter Gadamer's objection, since interpretations of the theoretical tools used to forge this theory may themselves be subject to other interpretations that vary across the cultural traditions of social interpreters.[10]

4.4 The Theory of Communicative Action

Gadamer's objections notwithstanding, Habermas has embarked on a a multi-volume statement of a comprehensive social theory centered on communicative action. In the first volume, Habermas [59] concentrates on the connection between the theory of universal pragmatics and the general theory of action descending from Weber [60] through Parsons [61] to Schutz [62] and Garfinkel [63]. His strategy is to align the various types of communication, their inherent truth claims, and their counterparts in rational action. Cognitive communication, in which the correspondence of terms to objects and events is at issue, has its rational action counterparts in *instrumental* and *strategic action.* These types of action are oriented toward success and are validated by *instrumental reason*, which reflects on the efficacy of plans as means to desired ends. Habermas ties interactive communication, in which claims to moral rightness and appropriateness are thematized, to *normatively regulated action*, in which the norms of a community and the social roles of actors become important constraints on the perceived appropriateness of actions. Finally, Habermas links expressive communication, in which the truthfulness of communicative actions are thematized, to *dramaturgical action*, which focuses upon the fact that actors respectively constitute a public for each other. Dramaturgical action attends to phenomena involving each actor's presentation of the self to others [63.5], to those aspects of the actor's subjectivity which he chooses to reveal to others and to those which he chooses to conceal. These revelations and concealments are, in turn, important factors which rational actors must assess when interpreting the actions of others and when planning their own.

5 Phenomenological Hermeneutics

Faced with the diversity of hermeneutics, and other continental philosophies including structuralism and phenomenology, Ricoeur strives for a grand synthesis in his phenomenological hermeneutics.[11] Ricoeur [64] argues that phenomenology and hermeneutics presuppose each other. The connection between hermeneutics and phenomenology traces to Heidegger who took the term "hermeneutics" from Dilthey to distinguish his own philosophical investigation of everyday being from Husserl's transcendental phenomenol-

[10] McCarthy [57] provides a short review of the debate and a discussion various problems in Habermas' position. McCarthy [58: ch. 3, 4] provides a more detail overview of the debate and a systematic rendition of Habermas' arguments. Ricoeur's proposed resolution of this debate is discussed below.

[11] For his interpretation of earlier hermeneuticists, see Ricoeur [63.6].

ogy, which tried to achieve objective knowledge by suspending concern for the subject's life-world. To capture knowledge of that world, Heidegger retained Husserl's notion of eidetic phenomenology, which assumes immediate registration of phenomena in a picture-like but uninterpreted manner. Like Heidegger, Ricoeur also follows Husserl to eidetic phenomenology, but like the late Heidegger and, particularly, Gadamer, Ricoeur recognizes the ontological basis of understanding in language. For Ricoeur, then, the subject's being is not identical with immediate experiences. So, instead of attempting a direct description of *Dasein* like Heidegger [4] and Merleau-Ponty [65, 66], Ricoeur sees the need for a hermeneutic theory of interpretation to uncover the underlying meaning constituting *Dasein.* Through its emphasis on the pre-linguistic, eidetic phenomenology supplies a means of distancing observation from linguistic descriptions that immediately call up preconceptions. This distanciation [66.5] is precisely what is required for interpretation to proceed. Since the task of uncovering the underlying objectivity cannot be achieved through the suspension of subjectivity, Ricoeur concludes that Husserl's project of transcendental phenomenology can only be realized through the application of a methodological hermeneutics to an eidetic phenomenology.

Ricoeur also argues that structuralism and hermeneutics can be complementary approaches to analyses of language, meaning, and cultural symbolism, for reasons similar to those he advanced for the complementarity of eidetic phenomenology and hermeneutics. *Structuralism* refers to a mode of inquiry which inventories elements of a system and notes the grammar of possible combinations. It is exemplified by Saussurean linguistics and Levi-Strauss' anthropology [67]. Ricoeur finds the value of structuralist analysis lies in its ability to catalogue phenomena and describe their possible combinations, but its weakness lies in its inability to provide anything more insightful than behavioral descriptions of closed systems. Nevertheless, the ability to generate structural descriptions complements the hermeneutic method, which interprets these descriptions by assigning functional roles to the phenomena.

In his treatment of psychoanalysis, particularly the interpretation of dreams, Ricoeur [68] shows the complexity involved in this hermeneutic task of assigning functional roles to words and symbols. The analyst must develop an interpretive system to analyze the dream-text and uncover the hidden meanings and desires behind its symbols, particularly those which have multiple senses (polysemy). Allowing for the possibility of multiple levels of coherent meaning, hermeneutics aims at ascertaining the deep meaning that may underlie the manifest or surface meaning. Ricoeur distinguishes two approaches for getting at the deeper meaning: a *demythologizing* one that recovers hidden meanings from symbols without destroying them (in the manner of the theologian Bultmann) and a *demystifying* one that destroys the symbols by showing that they present a false reality (in the manner of Marx, Nietzche, and Freud). The demythologizers treat the symbols as a window into a sacred reality they are trying to reach. But the demystifiers treat the same symbols as a false reality whose illusion must be exposed and dispelled so that a transformation of viewpoint may take place, as, for example, in Freud's discovery of infantile illusions in adult thinking. Thus, there are two opposing tendencies, a revolutionary and a conservative hermeneutics. Whereas the critical hermeneutics of Apel and Habermas falls within revolutionary demystification, the phenomenological hermeneutics of Ricoeur and the philosophical hermeneutics of Gadamer fall in the more conservative camp of the demythologizers.

Ricoeur [69] attempts a dialectical resolution of the Habermas-Gadamer debate, arguing that the hermeneutics of tradition and the critique of ideology require each other.

He denies the alleged antinomy between the ontology of tradition which limits possible meanings (Gadamer) and the eschatology of freedom which seeks to transcend these constraints (Habermas). If, as Gadamer believes, understanding should be conceived as the mediation between the interpreter's immediate horizons and his emerging horizon, then the interpreter must distance himself to some degree if he hopes to understand the text. That is, when confronted with a text, the interpreter must adopt a stance of critical self-understanding, not unlike the stance adopted in the critique of ideology. Hermeneutics thus incorporates a critique of ideology. Likewise, the critique of ideology incorporates tradition. The ideal of undistorted communication and the desire for emancipation do not begin with Habermas. They arise from a tradition – from the tradition of the Greek conception of "the good life", from the exodus, and from the resurrection. Thus the interests voiced by Gadamer and Habermas are, in Ricoeur's view, not incompatible. One is an interest in the reinterpretation of traditions from the past and the other is the utopian projection of a liberated humanity. Only when they are radically and artificially separated, argues Ricoeur, does each assume the character and tenor of ideology.

5.1 The Hermeneutic Arc: Ricoeur's Theory of Interpretation

Ricoeur's theory of interpretation [70] seeks a dialectical integration for Dilthey's dichotomy of explanation (*erklaren*) and existential understanding (*verstehen*). Ricoeur begins by distinguishing the fundamentally different interpretive paradigms for *discourse* (written text) and *dialogue* (hearing and speaking). Discourse differs from dialogue in being detached from the original circumstances which produced it, the intentions of the author are distant, the addressee is general rather than specific and ostensive references are absent. In a surprising move, Ricoeur extends his theory of interpretation to action, arguing that action evinces the same characteristics that set discourse apart from dialogue. A key idea in Ricoeur's view is that once objective meaning is released from the subjective intentions of the author, multiple acceptable interpretations become possible. Thus meaning is construed not just according to the author or agent's world-view but also according to its significance in the the reader's world-view.

Ricoeur's *hermeneutic arc* combines two distinct hermeneutics: one that moves from existential understanding to explanation and another that moves from explanation to existential understanding. In the first hermeneutic, subjective guessing is objectively validated. Here, understanding corresponds to a process of hypothesis formation, based on analogy, metaphor and other mechanisms for "divination." Hypothesis formation must not only propose senses for terms and readings for texts but also assign importance to parts and invoke hierarchical classificatory procedures. The wide range of hypothesis formation means that possible interpretations may be reached through many paths. Following Hirsch [26], explanation becomes a process of validating informed guesses. Validation proceeds through rational argument and debate based a model of judicial procedures in legal reasoning. It is therefore distinguished from verification which relies on logical proof. As Hirsch notes, this model may lead into a dilemma of "self-confirmability" when non-validatable hypotheses are proposed. Ricoeur escapes this dilemma by incorporating Popper's notion of "falsifiability" [71] into his methods for validation, which he applies to the internal coherence of an interpretation and the relative plausibility of competing interpretations.

In the second hermeneutic that moves from explanation to understanding, Ricoeur distinguishes two stances regarding the referential function of text: a subjective approach and a structuralist alternative. The subjective approach incrementally constructs the world that lies behind the text but must rely on the world-view of the interpreter for its pre-understanding. Although the constructed world-view may gradually approximate the author's as more text is interpreted, the interpreter's subjectivity cannot be fully overcome. In contrast, Ricoeur sees the structuralist approach as suspending reference to the world behind the text and focusing on a behavioral inventory of the interconnections of parts within the text. As noted earlier, the structural interpretation brings out both a surface and a depth interpretation. The *depth semantics* is not what the author intended to say but what the text is *about*, the non-ostensive reference of the text. Understanding requires an affinity between the reader and the *aboutness* of the text, that is, the kind of world opened up by the depth semantics of the text. Instead of imposing a fixed interpretation, the depth semantics channels thought in a certain direction. By suspending meaning and focusing on the formal algebra of the genres reflected in the text at various levels, the structural method gives rise to objectivity while capturing the subjectivity of both the author and the reader.

Like the other traditions, Ricoeur's hermeneutic arc can be interpreted as a bootstrapping process. Because he grounds the bootstrapping in an eidetic phenomenology, incorporates an internal referential model of the text constructed by the interpreter, and then begins interpretation with a structural analysis, Ricoeur's theory of interpretation may be easier to envision in computational terms. But the central bootstrapping locomotor in his theory is the alternation between forming hypotheses about meanings and validating those hypothesis through argument. This view resonates strongly with computational ideas about common-sense reasoning. Indeed, these ideas lead Ricoeur to identify metaphor as the main source of semantic innovation [71.4, 71.5], linguistic evolution, and therefore as major question for hermeneutics [71.6].[12]

6 Hermeneutics as Meta-Science

The hermeneutic tradition provides a basis for prescribing and criticizing the conduct of inquiry and the development of knowledge in the natural, social and cognitive sciences. Its representatives have figured prominently in debates concerning how valid knowledge can be acquired and whether there is a need for a separate methodology in the social sciences. Since AI is a new discipline, occupying a middle ground between the natural and social sciences, its researchers can benefit from knowledge of these debates. The choice of the appropriate methodology for AI research in many areas, like natural language processing, human problem solving, belief systems and action is unsettled. On one hand, the substantial contributions of logic, mathematics, engineering and the natural sciences, like physics, to AI seem to make functional model building the uncontested methodology. On the other hand, the subject matter is clearly linked to the concerns of linguistics and the social sciences, particularly anthropology and psychology, and thus the methodologies in those areas might be more appropriate for such research.

[12]For an excellent overview and comparison of the treatments of language and cognition found in phenomenological hermeneutics and in other non-hermeneutical traditions of philosophy, see Dallmayr [72].

6.1 Hermeneutics and the Social Sciences

Dilthey distinguished the cultural and social sciences (*Geisteswissenschaften*) from the natural sciences on the basis of their objects and the appropriate means for knowing them. The natural sciences concerned phenomena which, opaque to thought, could only be studied from the "outside" through observation of uniformities in their behavior and the construction of causal laws to explain those uniformities. In contrast, the human sciences had objects such as texts, verbal expressions and actions which could be investigated from the "inside" through an understanding of their authors' experiences and intentions. An interpretive or hermeneutic methodology could more reliably and intelligibly account for these objects by reconstructing the internal cognitive processes which motivated and gave meaning to each of them. The use of hypothetico-deductive methods employed in the natural sciences could only capture the external correlations among these objects at some high level of abstraction. Dilthey's arguments were embraced in the early 20th century by many social scientists, including the sociologist Max Weber [60], whose paradigmatic studies of social institutions interpreted human behavior as intentional action, structured by the agents' goals and belief.[13] However the physics model of the social sciences also persists and is currently manifested in such techniques as Skinnerean stimulus-response modeling of human behaviors and statistical content analysis, which determines the meaning of texts through frequency count of their words.

Contemporary hermeneuticists, such as Apel [73, 74], Habermas [44], and Ricoeur [75], strengthen Dilthey's distinction by noting that in the human sciences the subject of investigation and the investigator can communicate with each other. The equality suggests that an appropriate methodology will resemble discussions in which members in a community justify their actions. The tools of the natural sciences are simply incapable of representing the key concepts in such discussions, namely motivation, belief, and intention, and the complexity of their interactions. Intentional actions are embedded in groups of varying size and are constrained by (re-) created rules and norms – socio-cultural traditions. Because of the complexity of these intertwined and mutually-defining webs of relationships, scientific access to them is difficult and "uncertainty principles" abound – whether these involve the difficulties of isolating the object of study from its milieu or changes which communication between the investigator and the subject produces in the subject. These conditions reinforce the notion that cultural and social studies have the role of clarifying the beliefs, plans, motivations, and social roles that led cognitive agents to produce their texts and actions. The inquiry becomes a "dialogue" through which the inquirer comes to understand the tradition in which the author or agent is embedded, so that he may either consent to or repair the tradition, as Gadamer demands, or even reject it, as Habermas permits. Phases of understanding may be alternated with phases of validating knowledge, as Ricoeur's hermeneutic arc suggests, or of seeking explanations to opaque behaviors, as suggested in Apel's model of psychoanalysis. In any event, hermeneutic studies are inherently interactive and produce self-understanding. In this way, they extend the original mission of hermeneutics to mediate cultural traditions by correcting misreadings or distortions.

Logical positivists have nevertheless rejected the claims for a separate method for social and cultural sciences as groundless challenges to their own program of creating

[13] Weber considers means-ends rationality characteristic of intentional action in the modern age, but he also defines other types of intentionality.

a unified scientific method based on an unambiguous observation language [76]. Abel [77], Hempel and others argue that empathetic understanding and the attribution of rule following are psychological heuristics, unverifiable hunches or intuitions, based on personal experience. Although Abel concedes that they may be useful in setting up lawlike hypotheses for testing, he concludes that they are neither necessary nor sufficient to constitute a human science.

There are several rebuttals to these claims. First, methodological hermeneutics, which Dilthey initiated and which Betti [27] and Hirsch [26] continue, holds that an interpretation can be "objective" and "valid," if not verifiable, providing that the investigator resists temptations to make the text relevant for her own practical affairs. This strategy regards the text as an embodiment of the values of its time and suspends credibility regarding its truth and acceptability, according to present standards. But, knowledge of values expressed in other texts and records from the period are allowed to constrain the possible interpretations. Second, the idea of an interpretive or hermeneutic social science has received indirect support from ordinary language philosophy, an analytic that eschews the mentalism to which the logical positivists so strenuously object. The support comes from the sociologist Peter Winch [78], who generates recommendations for a social science on the basis of the later Wittgenstein's analysis [45] that particular word use and discourse patterns – "language games"— reflect and constitute activities in semi-institutionalized, functional areas of life – "life-forms." Winch consequently contends that the analysis of social actions (both verbal and non-verbal) has a necessarily holistic, situation oriented, interpretive character, rather than a generalizing, explanatory one.[14] Third, philosophical hermeneutics is not concerned with verifiable accounts and, as noted above, it denies the possibility of objective knowledge. Instead, it argues that only a person who stands in history, subject to the prejudices of his age, can hope to understand it. A valid understanding of an event, interaction or text is one that bridges history or socio-cultural differences to highlight the inquirer's situation.[15]

6.2 Hermeneutics and the Natural Sciences

Kuhn's influential *The Structure of Scientific Revolutions* [79] developed a hermeneutics of the natural sciences by portraying them as historically embedded, linguistically-mediated activities organized around paradigms which direct the conceptualization and investigation of the objects of their studies. *Scientific revolutions* occur when one paradigm replaces another and introduces a new set of theories, heuristics, exemplars and terms. The notion of a paradigm-centered scientific community consequently seems analogous to Gadamer's notion of a linguistically encoded social tradition. Kuhn [80] reports that his own development toward this idea began with his distress over Aristotle's theory of motion and the eventual discovery that Aristotle meant by "motion" something other

[14] "Understanding... is grasping the *point* or *meaning* of what is being done or said. This is a notion far removed from the world of statistics and causal laws: it is closer to the realm of discourse and to the internal relations that link the parts of ... a discourse" [78: 115].

[15] By this standard, Winch's recommendations are not hermeneutic, because they are based on the idea of ahistorical language games. They do not recognize that interpretation includes both "translation" and "application," that is, the mediation between the disintegrating and the emerging language-games, on one hand, and the revitalization of the past and its assimilation into the present life-form, on the other hand [73].

than what the word signified in Newtonian mechanics. This effort corresponds closely to a programmatic definition of hermeneutics as the study of human actions and texts with the view to discover their meanings in order to understand them, agree with them or even amend them [76].

Debates around Kuhn's thesis have spurred often grudging concessions that data, facts and law-like relations are theory-dependent rather than verifiable, coherent, and independent of the scientific theories in which they are embedded [81: 33]. Noting the inescapable theory-dependence of observational sentences and the incommensurabilities across paradigms, Feyerabend [82, 83] reaches the radical conclusion that no methodological standards can legitimately be applied. He therefore advocates a "methodological anarchism" that proceeds from the slogan, "in science, anything goes!" Feyerabend's doubts about the possibility of inter-paradigm communication closely resemble Gadamer's doubts regarding the accessibility of alien traditions.

Putnam [84: 113-119], however, argues that Feyerabend conflates concepts with conceptualization. According to Putnam, communication across paradigms does not require that the concepts be the same across paradigms, but only that members of one paradigm make their ideas *intelligible* to members of another paradigm. They can do so, if the apparati of conceptualization are the same across paradigms (language communities),nd the apparati are the same, because if they were not universal and *a priori*, an empirical experience would not be possible. Since the process of making ideas intelligible across paradigms would require reconstructions of the decisions leading to the concepts upon which a paradigm's theories rely, as well as reconstructions of the grounds of those concepts (and so on, recursively), inter-paradigmatic communication would accordingly require a critique not unlike that proposed by Apel and Habermas for exposing distortions in intersubjective communication. Additionally, as Apel [73] notes, justifications for scientific statements ultimately rely on ordinary language to establish procedural norms regarding evidence and argumentation. This common ground – the "communicative *a priori*"– reduces the room for arbitrariness and dogmatism.

6.3 The Notion of an Emancipatory Science

The hermeneutics tradition also provides the methodological starting point for Marx's critique of ideology, Freud's psychoanalysis and other studies which seek human emancipation by dissolving obsolete, restrictive and self-denying traditions and practices. Their initial strategy is to unmask the justifications given for these practices as distortions of the actors' true needs and the conditions of the situation. Yet, hermeneutic understanding will not reveal why the actors accept these justifications. In presenting psychoanalysis as the paradigmatic emancipatory science, Apel [42, 74] emphasizes that human beings cannot fully acknowledge their own motives or the intentions in their expressions. Consequently, empathy and introspection need to be supplemented by a quasi-naturalistic turn that applies the causal analysis of natural science to the actor's behavior. Any resulting explanations can then be fed back to the actor and appropriated as self-knowledge.

As mentioned earlier, Gadamer and Habermas debated the validity of rejecting past traditions, especially in regard to the critique of Western political and social institutions. Gadamer essentially considers this move incoherent and ungrounded, since it rejects the very tradition, including the value of rational, non-coerced consensus, that the investigator must accept to begin the explication. In response, Habermas [43] and Apel [73] claim that the preference for reason and understanding – the grounding for hermeneutics – is not just arbitrary or an inherited prejudice in the Western cultural tradition. Instead, they assert that a *communicative a priori* underlies all communication and entails that speech (and speech-like action) must be appropriate, as well as grammatical and sincere, to be meaningful. Since these validity claims imply a process for reaching agreement, the act of speaking itself commits the speakers to prefer reason.

7 Hermeneutics in AI

So far, few AI researchers have attempted to adopt AI techniques to hermeneutics and thereby develop computational models of interpretation. Hermeneutics, instead, appears a more fertile sponsor of arguments which doubt the possibility of successful cognitive models or of natural language processors. Nevertheless as AI interest in action and social discourse deepens, researchers will have to give hermeneutic insights a prominent place in thinking about the organization of these phenomena and the ability of AI models to capture their unfolding. Their modeling efforts are not likely to be very useful if they do not take into the account the variability of meaning according to the actors intentions and observer's perceptions and also the extent to which the modeling efforts themselves impute – rather than recognize – particular organization in the phenomena under study.

7.1 Analyzing the Affect Structure of Text

Lehnert, Alker, and Schneider [85, 5] present a bottom-up model for extracting the affective structure of a text. Their "computational hermeneutics" builds from Lehnert's earlier work on "plot units" [86, 87]. *Plot units* provide a unvalidated but nevertheless interesting vocabulary for designating affective relationships and their combinations. In this research, they are used to describe many emotional consequences for participants in events and actions. Working within "conceptual dependency" theory [88], Lehnert identified various combinations of plot units for use in summarizing narrative texts. These "story molecules" relate changes in actors' affects to successes and failures in the resolutions of problems involving them. In their work, Lehnert, Alker, and Schneider manually reduced passages from Toynbee's retelling of events leading up to Christ's crucifixion into a large number of these molecules. The molecules were interrelated through the actors involved and by virtue of some molecules being antecedent conditions for others. After the input of these manual reductions, the central subgraph of the plot structure was computationally extracted, using a program for finding the most strategic and highly connected molecules. This central subgraph was labeled the "essential" Jesus story.

After studying this affective core, Alker, Lehnert and Schneider concluded that the Jesus story involves an ironic victory born from adversity and conforms to a well-known genre, the romance of self-transcendence. Their method resembles classical hermeneutics in seeking to uncover the essential structure of text based on systematic linkages between the parts and the whole and in emphasizing the use of explicit rules for objective interpretation. However, their willingness to tolerate multiple interpretations and their structuralist orientation also aligns them with phenomenological hermeneutics. Alker, Lehnert and Schneider suggest that the Jesus story has been emotively potent because it provides a step-by-step account of affective change in self-transcendence and thus can open its readers to the experience of this process. In its present form, however, this work does not implement a bootstrapping process, even though the theme of self-transcendence presupposes a mechanism capable of consciously-directly bootstrapping.

7.2 What Does it Mean To Understand Natural Language?

Winograd [2] uses insights primarily from philosophical hermeneutics to sketch a new approach to natural language understanding. He intends to overcome the pitfalls of earlier approaches that succumbed to the phenomenological critique advanced by Dreyfus [1]. Focusing on the theory of meaning, Winograd argues that previous efforts, including his own SHRDLU [89], fell into the trap of "objectivism," or the misplaced belief that the contents of a theory or model correspond directly to reality (the Correspondence Theory of Truth).[16] He adds that the deductive nature of the formalisms used by AI researchers forced them to adopt an objectivist position, but that these formalisms failed to account for the informal, phenomenological knowledge or experience that an understander deploys when interpreting utterances. Hermeneuticists identify this problem as the historicity of understanding or the role of background knowledge in mediating understanding.

Moreover, these deductive formalisms are subject to Heidegger's ontological critique of Husserl. Their failure to address the fundamental ontology of language typified by the conversational situation leads to an inability to account for the role of context in speaker/hearer identification of the intended meanings of utterances [2: 215-219]. Thus, Winograd supports the Heideggerian critique with arguments and examples drawn from ordinary language philosophy [53, 55, 91, 92]. In a vein reminiscent of Gadamer, he argues that making sense of a statement requires knowing how it is intended to answer (implicit or explicit) questions posed by the conversational context. He concludes that deductive logic can account for only a small fraction of human reasoning, and therefore new advances in natural language understanding require "a calculus of natural reasoning" [2: 219].

Winograd proposes *Knowledge Representation Language* (KRL) [93] as a starting point for an alternative approach. KRL's reasoning based on limited computational resources captures Heidegger's thesis of the finititude of *Dasein* and also echoes Simon's notion of "bounded rationality" in the theory of decision-making [94]. For Winograd, effective reasoning strategies under limited or variable computational resources provide a "natural reasoning," which, although formally incomplete, can account for more of everyday natural language usage than can the small fraction that fits the mold of a complete deductive logic [2: 219]. Moreover, this approach must have the ability to deal with partial or imprecise information if it is to work at all. Winograd proposes a control structure that uses matching of the current processing context to trigger actions appropriate for the situation. This view of situated action, in which situational responses are unreflective, resembles the concept of "thrownness" as developed in Heidegger's notion

[16] Prior [90] provides a concise overview of the Correspondence Theory of Truth, which holds that the structure of theoretical knowledge corresponds to reality.

of "the projected world." The combination of situated action as a control structure and resource-limited reasoning grounded in common-sense, stereotype-based reasoning [2: 220, 235] resonates with recent work on analogy [95, 96, 97: ch. 12], precedential reasoning [98], and metaphor [99, 16]. At its core, KRL also incorporates a notion of bootstrapping similar to the one found in the various hermeneutic traditions, particularly in the works of Heidegger and Gadamer.

Winograd argues that *spurious reification*, or misplaced concreteness, has plagued earlier efforts to develop a formalisms for representing natural language. Spurious reification occurs when a competence is imputed to an understander not because the understander actually employs the specified competence in performance, but because the observer classifies performances as instances of a particular competence, and then mistakenly imputes the competence to the understander. Instead of building from domain-level concepts and structures, Winograd attempts to avoid spurious reification by constructing formal representations based on ontological considerations borrowed from methodological hermeneutics [2: 227]. Since no substantial AI project has been attempted using KRL, the ideas that its designers hoped to capture remain more theoretical than practical.

In discussing hermeneutics, Winograd not only proposes a new research program for AI but also problematizes the philosophical basis of current natural language research. Fundamental assumptions and philosophical orientations underlying research must now be explicitly analyzed and justified. In rejecting "objectivism," Winograd advocates a "subjectivist" hermeneutical position that builds from Maturana's [100] notion of the nervous system as "structure determined," plastic and closed. According to this model, activities outside the system (stimuli) perturbate the structure of the system and these perturbations in turn lead to "patterns of activity that are different from those that would have happened with different perturbations." Winograd's parallel notion of understanding posits a closed system in which pre-understanding evolves through acts of interpretation. As in Heidegger's hermeneutic circle, the possible horizons that can be understood are constrained by the historically determined structure of pre-understanding or set of stored schemas [2: 223-224]. Understanding is open to the environment, only within this range. Unlike Heidegger, who recognized the importance of the environment but failed to analyze it, Winograd is led to the analysis of the environment by several influences. These include Garfinkel's [63] ethnomethodology, which emphasizes social context, Searle's focus on speech as social action, and Lakatos' [101] argument that even in mathematics the meanings of terms are contingent on outside context. Winograd [2: 231-33] grounds his theory of meaning in terms of social action, and so takes a position close to critical hermeneutics, between relativism and objectivism.

Stimulated in part by Winograd [2], Bateman [6, 7] examines the consequences of Heidegger's existential phenomenology and agrees with Dreyfus [1] that this philosophy denies the possibility of modeling thought and action using the specific formalizations proposed by the functionalist paradigm of cognitive science. Bateman says these formalisms are based on the "ontological assumption" of an interpreter who follows rules in acting upon a mental representation of a situation. Heidegger's notion of "being-in-the-world" which includes both situatedness and understanding as ontological modes, precludes the subject-object dichotomy in this assumption. Since one is always in a situation, and its structure and significance are determined by its relevance to one's plans and purposes, no context-free representation is possible.

Bateman, however, does not dismiss the possibility of a functionalist paradigm for cognitive science. He wants instead to ground it on the later Heidegger's idea of language, which, according to Bateman, seeks to make intelligible the experience of "being-in-the-world" as it is for "anyone," that is, for a generalized subject or member of a language community. As a collective artifact, a language is considered to encode partially the history of the language community through both the admissible and inadmissible combination (association) of words and phrases. The resulting connotational structure captures a kind of collective background knowledge and imposes *a priori* constraints on the actions of individuals who contemplate actions in terms of the language. In Halliday's "systemic grammar" [102], there is the notion of a "social semiotic" that acknowledges that a group's culture can restrict the possible meanings of utterances through constraints on possible ways of acting in situations. Bateman considers this orientation compatible with the hermeneutic view, and believes that "systemic grammar," with appropriate revisions, can provide an adequate theoretical framework for natural language understanding. Yet despite this openness to social constraints, Bateman does not consider hermeneuticists who came after Heidegger, most notably Gadamer and Habermas.

7.3 The Foundations of Understanding

In a more recent work, Winograd and Flores [3] draw upon philosophical hermeneutics and Maturana's [103] work on the biology of cognition to deny the possibility of the construction of intelligent computers is impossible, as is computer understanding of natural language. They argue that to the extent Heidegger and Gadamer make a persuasive case that certain fundamental features of human existence are fundamental, the quest for intelligent machinery is quixotic. These concepts include "thrownness", "blindness", and "breakdown." "Thrownness" denotes that people are thrown into the situations of everyday life and rarely have time to reflect upon alternative courses of action. They cannot be impartial, detached observers of the world in which they live, but they must decide and act using heuristics which they have as part of their effective histories. Although these heuristics enable some action possibilities, the same heuristics also "blind" people to other action possibilities that might have predominated had their effective-histories been different. When faced with situations where their effective-histories fail to provide an adequate guide for action and also "blind" them to those actions that support their purposes, people experience a kind of "breakdown" which can take the form of severe depression, anomie, and/or irrational action.

If an expert system is designed to present a user with possible courses of action in particular situations, the concepts of "thrownness," "blindness", and "breakdown" also come into play. Although expert systems may operate successfully in well-understood, constrained domains, expert systems in complex domains may be "thrown" into situations where they cannot evaluate all possible actions and they consequently "break down." Systems targeted at complex domains must therefore rely on heuristic rules but these may "blind" the program to more propitious courses of action. Winograd and Flores add that the expert-system programmer introduces his own "blindness" or preconceptions into the program. Because of these difficulties, Winograd and Flores recommend reformulation of the goals of artificial intelligence. Instead of directing efforts toward the putatively impossible goal of creating machines that can understand, programs should be designed to serve as tools for enhancing the quality of life. This could be done by recognizing the role of such programs in the web of conversations (speech acts) that constitute social existence, by attempting to minimize the "blindness" they engender and by anticipating the range of their potential "breakdowns."

Winograd and Flores present a reasoned critique of two specific categories of AI research. The first comprises AI approaches that incorporate rigidly fixed means of interpretation, such as much work in knowledge-based systems. The second category includes those approaches which proceed from the dualist presumption that truth, meaning, and reference are established by means of a correspondence between entities in the world and entities in the mind (the Correspondence Theory of Truth) rather than in the everyday discourse of intelligent agents. Although they acknowledge that learning approaches might eventually be able to address the criticisms they raise, they do not expect progress in learning during the near term. Thus, their work amounts to a critique of the tractability of the "hard AI" project. As such, it constitutes a continuation of the critique AI begun by Dreyfus [1] but differs in that it comes from within AI and is argued in more computational terms.

However, Winograd and Flores fail to demonstrate convincingly demonstrate that computer understanding exceeds the range of the possible. They only demonstrate that the goal is much more difficult than many people, including many AI practitioners, may have thought. Unfortunately, Winograd and Flores unfairly characterize as "blind" those AI approaches which come closest to overcoming their objections, such as Winston's [95, 104, 97] approach to learning and reasoning by analogy. Winograd and Flores misconstrue Winston's approach as capable of producing results only because it operates in a microworld with primitives fixed in advance by the implementors. Although this criticism may be leveled fairly at many AI programs, Winston's program is in principle not so limited, precisely because it is not based on domain-specific primitives. Indeed, Winston's program is general enough to perform well in any domain because it processes linguistically derived data according to the data's form rather than its specific content. Moreover, because it learns rules on the basis of its experience (its effective history) over which it can draw analogies, Winston's program represents a first computational approximation of the basic hermeneutic notion of a pre-understanding grounded in effective-history.

7.4 Grounding Meaning in Eidetic Phenomenology

Mallery and Duffy [8] present a computational model of *semantic perception* – the process of mapping from a syntactic representation into a semantic representation. Some computational and non-computational linguists [105, 106, 88] advocate determining equivalent meanings (paraphrases) through the reduction of different surface forms to a canonicalized semantic form, a combination of semantic universals (*e.g.*, "conceptual dependency" primitives). Mallery and Duffy reject this view on the grounds that most meaning equivalences must be determined in accordance with the specific linguistic histories of individual language users, or at least linguistic communities, and the intentional context of the utterance. Their alternative is *lexical-interpretive semantics*, an approach to natural-language semantics that constructs semantic representations from canonical grammatical relations and the original lexical items. On this view, semantic representations are canonicalized only syntactically, not semantically or pragmatically. Instead of relying on static equivalences determined in advance, lexical-interpretive semantics requires meaning equivalences to be determined at their time of use, reference time. To meet this requirement, Mallery and Duffy introduce the concept of a *meaning congruence class*, the set of syntactically normalized lexical realizations conforming to the linguistic experience of specific language users and satisfying their utterance-specific intentions. Meaning equivalences are then given by the meaning congruence classes to which utterances belong. In sum, lexical-interpretive semantics differs from approaches relying on semantic universals because meaning equivalences are determined dynamically at reference time for specific language users with individual histories rather than statically in advance for an idealized language user with an unspecific background knowledge but who is "semantically and pragmatically competent."

In principle, lexical-interpretive semantics can avoid the distorting pitfalls incurred by exclusive reliance on a static analysis of meaning equivalence precisely because it can select equivalences from synonym or paraphrase equivalence classes on the basis of dynamically changing, intentional contexts of language use. The major assumption underlying lexical-interpretive semantics is that meaning equivalences arise because alternative lexical realizations accomplish sufficiently similar speaker goals to allow substitution. A practical argument for dynamically determining meaning equivalences is the intractability of a sufficiently detailed and nuanced static analysis which is capable of capturing subtle differences in speaker goals as expressed surface sentences. This follows from the need to predict in advance all potential utterance situations and combinations of language-user effective-histories. Although semantic canonicalization on the basis of a general "semantic and pragmatic competence" renders static analyses of language-user combinations tractable by fiat, it also reduces nuances so dramatically that intentional analysis plays a drastically diminished role.

Lexical-interpretive semantics is hermeneutic because it emphasizes interpretation based on the individual effective-history of language users and the specific intentional structure of communicative situations. By virtue of its emphasis on innovation in language and polysemy, lexical-interpretive semantics is perhaps most closely aligned with the phenomenological hermeneutics of Ricoeur [64]. Interpretation builds from an *eidetic* level of representation, the syntactically normalized semantic representation. The determination of meaning congruence classes becomes an early level of a more general and open-ended hermeneutic interpretation. Stimulated by recent discussions of perception by Feigenbaum and Simon [107], Mallery and Duffy consider semantic perception to be a process of mapping from sense-data, in this case natural language sentences, to a semantic representation. But instead of providing an account of perception suited to a theory of meaning based on semantic universals like Feigenbaum and Simon, Mallery and Duffy provide one suited to a hermeneutic theory of meaning.

Mallery and Duffy have implemented this theory, up to the level of eidetic representation, in the RELATUS Natural Language System [108]. Although they share some of the hermeneutically oriented views and concerns articulated in Winograd [2] and Bateman [6, 7], their implementation allows more concrete specification and testing of their theory, which currently focuses on earlier processing levels. For example, Mallery and Duffy [8] have proposed constraint-interpreting reference [109] as a model that conforms to lexical-interpretive semantics, just as discrimination nets are well-suited to approaches relying on semantic primitives [110, 111, 112, 107]. They ground this choice both in the available experimental psycholinguistic evidence and in the desirable computational properties of reference based on constraint interpretation. These properties include maximizing monotonicity (minimizing backtracking) in the syntactic processing that precedes reference and optimizing subgraph isomorphism (search) as it arises in reference and in other reasoning operations – particularly common-sense reasoning grounded in analogy.

8 Conclusions

This article has presented hermeneutics primarily as a philosophy of understanding rather than as a set of technologies for interpretation in specific domains. As such, the hermeneutic tradition seems able to speak to AI researchers in two distinct ways. First, hermeneutics provides some basis for arguing against the feasibility of the AI project, at least under its present dispensation. Whether represented by Dilthey's idea of empathetic understanding or Heidegger's idea of situated understanding, hermeneutics seems to have discovered a quality in the human situation that is vital for knowledge of others and oneself but has not yet been simulated mechanically. Because these doubts are generated from a ongoing intellectual tradition and because they refine some fairly common intuitions, they cannot easily be dismissed as "irrational technological pessimism." On the other hand, these doubts should stimulate attempts by AI researchers to overcome them, as were some of the doubts raised by Dreyfus (1972). At the very least, then, the insights of the various hermeneutical camps can be expected to receive increasing attention in the artificial intelligence community.

Second, hermeneutics can suggest constraints, orientations and even criteria in the design of AI systems that are intended either to understand natural language or to represent knowledge of the social world. The lessons of this tradition are, however, equivocal. Dilthey, Heidegger, Gadamer, Habermas, Ricoeur and others provide very different notions of what constitutes understanding and its grounding. Nevertheless, researchers who are aware of these debates might be more cognizant of the choices they make in their own designs. As a consequence, systems would not merely illustrate isolated and perhaps idiosyncratic theories about linguistic phenomena but would begin to support (or deny) major philosophical positions in ontology, epistemology and philosophy of mind. But the generally pre-computational nature of contemporary hermeneutics calls for specific formulations that can be tested computationally. Computational experimentation, an empirical philosophy, can then feed back into the reformulation and refinement of ideas about both hermeneutics and AI.

9 Acknowledgements

This paper was improved by comments from Phil Agre, Hayward R. Alker, Jr., Doug Appelt, Hubert Dreyfus, David Kirsh, William J. Rapaport, Tom Reinhardt, and Daniel Schneider. This research was conducted 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. The views and conclusions are those of the authors, not the Massachusetts Institute of Technology or the United States Government. Any errors or shortcomings are the responsibility of the authors.

10 References

[1] H. Dreyfus, *What Computers Can't Do: A Critique of Artificial Reason*, Freeman, San Francisco, 1972. 2nd edition with a new preface was published in 1979.

[2] T. Winograd, "What Does It Mean to Understand Natural Language," *Cognitive Science* 4, 209-241 (1980).

[3] T. Winograd and F. Flores, *Understanding Computers and Cognition: A New Foundation for Design*, Ablex, Norwood, NJ, 1986.

[4] M. Heidegger, *Being and Time*, J. Macqarrie and E. Robinson, trans, Harper and Row, New York, 1962. Originally published as *Sein und Zeit*, Neomarius Verlag, Tubingen, 1927.

[5] H.R. Alker Jr., W.G. Lehnert, and D.K. Schneider, "Two Reinterpretations of Toynbee's Jesus: Explorations in Computational Hermeneutics," in G. Tonfoni, ed, *Artificial Intelligence and Text Understanding, Quaderni di Ricerca Linguistica* 6, 49-94 (1985).

[6] J.A. Bateman, "Cognitive Science Meets Existential Phenomenology: Collapse or Synthesis?," Working Paper No. 139, Department of Artificial Intelligence, University of Edinburgh, Edinburgh, April, 1983.

[7] J.A. Bateman, "The Role of Language in The Maintenance of Intersubjectivity: A Computational Investigation," in G. N. Gilbert and C. Heath, *Social Action And Artificial Intelligence*, Grower, Brookfield, VT, 1985, 40-81.

[8] J.C. Mallery and G. Duffy, "A Computational Model of Sematic Perception," AI Memo No. 799, Artificial Intelligence Laboratory, MIT, Cambridge, May 1986.

[9] W. Stegmuller, "The So-called Circle of Understanding," in W. Stegmuller, *Collected Papers on Epistemology, Philosophy of Science and History of Philosophy*, vol. 3, Reidel, Dordrecht, 1977.

[10] D.B. Lenat, "AM: Discovery in Mathematics as Heuristic Search," in R. Davis and D.B. Lenat, *Knowledge-Based Systems in Artificial Intelligence*, McGraw-Hill, New York, 1982, 1-227.

[11] D.B. Lenat, "Eurisko: A Program That Learns New Heuristics and Domain Concepts: The Nature of Heuristics III: Program Design and Results," *Artificial Intelligence* 21, 61-98 (1983).

[12] K. Haase, *ARLO: The Implementation of a Language for Describing Representation Languages*, Bachelor's Thesis, Department of Philosophy and Linguistics, MIT, 1984.

[13] J. Piaget, *The Origins of Intelligence in Children*, M. Cook, trans, W.W. Norton, New York, 1952.

[14] J. Piaget, *Genetic Epistemology*, Columbia University Press, New York, 1970.

[15] G.L. Drescher, "Genetic AI: Translating Piaget Into LISP," AI Memo No. 890, Artificial Intelligence Laboratory, MIT, February 1986.

[16] M. Minsky, *The Society of Mind*, Simon and Schuster, New York, 1986.

[17] R. Palmer, *Hermeneutics: Interpretation Theory in Schleiermacher, Dilthey, Heidegger, and Gadamer*, Northwestern University Press, Evanston, IL, 1969.

[18] J. Bleicher, *Contemporary Hermeneutics: Hermeneutics as Method, Philosophy, and Critique*, Routledge & Kegan Paul, London, 1980.

[19] B. Smalley, *The Study of the Bible in The Middle Ages*, 2d edition, Blackwell, Oxford, l952.

[20] F. Schleiermacher, *Hermeneutik*, H. Kimmerle, ed, Carl Winter Universitatsverlag, Heidelberg, l959.

[21] J.B. Thompson, *Critical Hermeneutics: A Study in the Thought of Paul Ricoeur and Jurgen Habermas*, Cambridge University Press, Cambridge, 1981.

[22] W. Dilthey, "The Rise of Hermeneutics," T. Hall, trans, in P. Connerton, ed, *Critical Sociology: Selected Readings*, Penguin, Harmondsworth, UK, 1976, 104-116. Excerpted from W. Dilthey, "Die Entstehung der Hermeneutik," 1900, in W. Dilthey, *Gesammelte Schriften*, B.G Teubner, Leipzig and Berlin, 1923, 317-20, 323-31.

[23] W. Dilthey, *Selected Writings*, H.P. Rickman, ed, Cambridge University Press, Cambridge, 1976.

[24] M. Ermarth, *Wilhelm Dilthey: The Critique of Historical Reason*, University of Chicago Press, Chicago, 1978.

[25] T. Plantinga, *Historical Understanding in the Thought of Wilhelm Dilthey*, University of Toronto Press, Toronto, 1980.

[26] E.D. Hirsch Jr., *Validity in Interpretation*, Yale University Press, New Haven, l967.

[27] E. Betti, "Hermeneutics as The General Methodology of The *Geisteswissenschaften*," in [18: 51-94]. Originally published as "Die Hermeneutik als allgemeine Method der Geisteswissenschaften" Mohr, Tubingen, 1962.

[28] R. Schmitt, "Phenomenology," in [113: vol. 5-6, 135-151].

[29] R.M. Zaner, *The Way of Phenomenology: Criticism as a Philosophical Discipline*, Pegasus, New York, 1970.

[30] P. Ricoeur, *Husserl: An Analysis of His Phenomenology*, E.G. Ballard and L.E. Embree, trans, Northwestern University Press, Evanston, IL, 1967.

[31] E. Husserl, *Ideas: General Introduction to Pure Phenomenology*, W.R.B. Gibson, trans, George Allen and Unwin, London, 1931. First published in 1913.

[32] G.W.F. Hegel, *The Philosophy of Mind*, Part three of *The Encyclopedia of the Philosophical Sciences*, W. Wallace, trans, Oxford University Press, Oxford, 1971. First published in 1830.

[33] G.W.F. Hegel, *The Science of Logic*, Part two of *The Encyclopedia of the Philosophical Sciences*, W. Wallace, trans, Oxford University Press, Oxford, 1975. First published in 1830.

P. Singer, *Hegel*, Oxford University Press, Oxford, 1983.

[34] M. Green, "Martin Heidegger," in [113: vol. 7-8, 457-465].

[35] G. Steiner, *Martin Heidegger*, Penguin, New York, 1980.

[36] M. Murray, *Heidegger and Modern Philosophy: Critical Essays*, Yale University Press, New Haven, 1978.

[37] H. Gadamer, *Truth and Method*, Continuum, New York, 1975. Originally published as *Wahrheit und Methode*, Tubingen, 1960.

[37.4] H. Gadamer, "Man and Language," in [114: 59-68].

[37.5] E. Sapir, *Selected Writings of Edward Sapir*, U. of California Press, Berkeley, 1947.

[37.6] B. Whorf,*Language, Thought and Reality*, M.I.T. Press, Cambridge, MA, 1967.

[38] J. Habermas, "A Review of Gadamer's *Truth and Method*," in [115: 335-363]. Originally published in *Zur Logik der Sozialwissenschaften*, Suhrkamp Verlag, Frankfurt am Main, 1970.

[39] D.E. Linge, "Editor's introduction," in [114: xi-lviii].

[40] H. Gadamer, "Hegel and Heidegger," in [116: 100-116].

[41] H. Gadamer, "The Idea of Hegel's Logic," in [116: 75-99].

[42] K. Apel, "Scientistics, Hermeneutics and The Critique of Ideology: Outline of a Theory of Science from a Cognitive-Anthropological Standpoint," in [73], 46-76.

[43] J. Habermas, "Knowledge and Human Interest," *Inquiry* 9, 285-300, (1966).

[44] J. Habermas, *Knowledge and Human Interests*, J.J. Shapiro, trans, Heinemann, London, 1972. Originally published in 1968.

[45] L. Wittgenstein, *Philosophical Investigations*, 3d edition, MacMillan, New York, 1968. Earlier edition published in 1953.

[46] N. Chomsky, *Syntactic Structures*, Mouton, The Hague, 1957.

[47] N. Chomsky, *Aspects of The Theory of Syntax*, MIT Press, Cambridge, 1965.

[48] N. Chomsky, *Lectures on Government and Binding,*, Foris, Dordrecht, 1981.

[49] K. Apel, "Noam Chomsky's Theory of Language and Contemporary Philosophy: A Case Study in the Philosophy of Science," in [73], 180-224.

[50] J. Habermas, "What is Universal Pragmatics?" in J. Habermas, *Communication and the Evolution of Society*, T. McCarthy, trans, Beacon Press, Boston, 1979, 1-68. First published in German in 1976.

[51] J. Habermas, "Some Distinctions in Universal Pragmatics, *Theory and Society* 3, 155-167, (1976).

[52] J.B. Thompson, "Universal Pragmatics," in [117: 116-133].

[53] J.L. Austin, *How To Do Things with Words*, Harvard University Press, Cambridge, 1962.

[54] P.H. Grice, "Logic and Conversation," in P. Cole, and J.L. Morgan, eds, *Studies in Syntax*, 3, New York: Academic Press, 1975, 41-58.

[55] J.R. Searle, *Speech Acts*, Cambridge University Press, Cambridge, 1970.

[56] H. Gadamer, "On The Scope And Function of Hermeneutical Reflection," in [114: 18-43].

[57] T. McCarthy, "Rationality and Relativism: Habermas' 'Overcoming' of Hermeneutics," in [117: 57-78].

[58] T. McCarthy, *The Critical Theory of Jurgen Habermas*, MIT Press, Cambridge, 1978.

[59] J. Habermas, *The Theory of Communicative Action*, vol. 1, *Reason and the Rationalization of Society*, T. McCarthy, trans, Beacon Press, Boston, 1981. German edition published in 1981.

[60] M. Weber, *The Methodology of the Social Sciences*, E. Shils and H. Finch, eds and trans, Free Press, Glencoe, IL, 1949.

[61] T. Parsons, *The Structure of Social Action*, McGraw-Hill, New York, 1937.

[62] A. Schutz, *The Phenomenology of a Social World*, Northwestern University Press, Evanston, IL, 1967.

[63] H. Garfinkel, "What is Ethnomethodology?" in [115: 240-261]. Originally published in H. Garfinkel, *Studies in Ethnomethodology*, Prentice-Hall, Englewood Cliffs, NJ, 1967.

[63.5] E. Goffman, *The Presentation of Self in Everyday Life*, Doubleday, New York, 1959.

[63.6] P. Ricoeur, "The Task of Hermeneutics," *Philosophy Today* 17, (1973), D. Pellauer, trans. Reprinted in [36: 141-160, 118: 43-62].

[64] P. Ricoeur, "Phenomenology and Hermeneutics," Translated and reprinted in [118: 101-128]. Originally published as "Phenomenologie et Hermeneutique," *Phanomenologische Forschungen*, vol. 1, E.W. Orth, ed, Karl Alber, Freiberg, 1975, 31-77.

[65] M. Merleau-Ponty, *Phenomenology of Perception*, C. Smith, trans, Routledge & Kegan Paul, London, 1962. Originally published as *Phenomenologie de la Perception*, Paris, 1945.

[66] F.A. Olafson, "Maurice Merleau-Ponty," in [113: vol., 5-6, 279-282].

[66.5] P. Ricoeur, "The Hermeneutical Function of Distanciation," *Philosophy Today* 17, (1973), 129-43. Reprinted in [36: 131-144].

[67] C. Levi-Strauss, *Structural Anthropology*, C. Jacobson and B.G. Schoepf, trans, Penguin, Harmondsworth, 1968.

[68] P. Ricoeur, *Freud and Philosophy: An Essay on Interpretation*, D. Savage, trans, Yale University Press, New Haven, 1970.

[69] P. Ricoeur, "Hermeneutics and the Critique of Ideology," in [118: 63-100]. Originally published as "Hermeneutique et critique des ideologies," in E. Castelli, ed, *Demythisation et ideologie*, Aubier Montaigne, Paris, 1973, 25-64.

[70] P. Ricoeur, "The Model of Text: Meaningful Action Considered as Text," *Social Research* 38, 529-62 (1971). Reprinted in [118].

[71] K. Popper, *The Logic of Scientific Discovery*, Basic Books, New York, 1959.

[71.4] P. Ricoeur, "Creativity in Language," *Philosophy Today* 17, (1973), 97-111.

[71.5] P. Ricoeur, *The Rule of Metaphor: Multi-Disciplinary Studies of the Creation of Meaning in Language*, R. Czerny, *et al.*, trans, University of Toronto Press, Toronto, 1977. Originally published as *La Metaphore vive*, edition du Seuil, Paris, 1975.

[71.6] P. Ricoeur, "Metaphor and the Main Problem of Hermeneutics," *New Literary History* 6, (1974-75), 95-110. Reprinted in C.E. Reagan and D. Stewart, eds, *The Philosophy of Paul Ricoeur: An Anthology of His Work*, Beacon Press, Boston, 1978, 134-148.

[72] F.R. Dallmayr, *Language and Politics: Why Does Language Matter to Political Philosophy?*," University of Notre Dame Press, Notre Dame, IL, 1984.

[73] K. Apel, *Towards A Transformation Of Philosophy*, G. Adey and D. Frisby, trans, Routledge & Kegan Paul, London, 1980. Originally published in *Transformation der Philosophie*, Suhrkamp Verlag, Frankfurt am Main, 1972, 1973.

[74] K. Apel, *Understanding and Explanation*, G. Warnke, trans, MIT Press, Cambridge, l984. Originally published as *Die Erklaren-Verstehen-Kontroverse in Tranzendental-Pragmatischer Sicht*, Suhrkamp, Frankfurt am Main, 1979.

[75] P. Ricoeur, *Main Trends in Philosophy*, Holmes and Meier, New York, 1979. Reprinted from *Main Trends in The Social and Human Sciences – Part II*,, UNESCO, New York,

1978.

[76] G. Radnitzky, *Continental Schools of Metascience: The Metascience of the Human Sciences based upon the "Hermeneutic-Dialectic" school of philosophy*, vol. 2 of *Contemporary Schools of Metascience*, Scandinavian University Books, Goteborg, Sweden, 1968.

[77] T. Abel, "The Operation Called *Verstehen*," in [115: 81-92]. Originally published in *The American Journal of Sociology* 54, 1948, 211-218.

[78] P. Winch, *The Idea of a Social Science and its Relation to Philosophy*, Routledge & Kegan Paul, London, 1958.

[79] T.S. Kuhn, *The Structure of Scientific Revolutions* University of Chicago Press, Chicago, 1962.

[80] T.S. Kuhn, *The Essential Tension: Selected Studies in Scientific Tradition and Change*, University of Chicago Press, Chicago, 1977.

[81] R.J. Bernstein, *Beyond Objectivism and Relativism: Science, Hermeneutics, and Praxis*, University of Pennsylvania Press, Philadelphia, 1983.

[82] P. Feyerabend, "Consolations for the Specialist," in I. Lakatos and A. Musgrave, eds, *Criticism and the Growth of Knowledge* Cambridge, University Press, Cambridge, 1970, 197-230.

[83] P. Feyerabend, *Against Method* Verso, London, 1978.

[84] H. Putnam, *Reason, Truth and History*, Cambridge University Press, Cambridge, 1981.

[85] W.C. Lehnert, H.R. Alker Jr. and D.K. Schneider, "The Heroic Jesus: The Affective Plot Structure of Toynbee's Christus Patiens," in *Proceedings of The Sixth International Conference on Computers and the Humanities*, S.K. Burton and D.D. Short, eds, Computer Science Press, Rockville, MD, 1983, 358-367.

[86] W.C. Lehnert, "Plot Units and Narrative Summarization," *Cognitive Science* 4, 293-331 (1981).

[87] W.C. Lehnert, "Plot Units: A Narrative Summarization Strategy," in [119: 375-414].

[88] R.C. Schank and R. Abelson, *Scripts, Plans, Goals, and Understanding*, Lawrence Erlbaum Associates, Hillsdale, NJ, 1977.

[89] T. Winograd, *Understanding Natural Language*, Academic Press, New York, 1972.

[90] A. N. Prior, "Correspondence Theory of Truth," in [113: vol. 1-2, 223-232].

[91] J.R. Searle, "A Taxonomy of Illocutionary Acts," in K. Gunderson, ed, *Language And Knowledge: Minnesota Studies In Philosophy Of Science*, 11, University of Minnesota Press, Minneapolis, 1975, 344-369.

[92] J.R. Searle, "The Intentionality of Intention and Action," *Cognitive Science* 4, 47-70 (1980).

[93] D. G. Bobrow, and T. Winograd, "An Overview of KRL, A Knowledge Representation Language," *Cognitive Science* 1, 3-46 (1977).

[94] H.A. Simon, "Rational Decision Making in Business Organizations," *American Economic Review* 69, 493-513 (1979).

[95] P.H. Winston, "Learning and Reasoning by Analogy," *Communications of the ACM* 23, December, (1980).

[96] J.G. Carbonell, "Learning by Analogy: Formulating and Generalizing Plans From Past Experience," in R.S. Michalski, J.G. Carbonell and T.M. Mitchell, *Machine Learning: An Artificial Intelligence Approach*, Tioga, Palo Alto, CA, 1983, 137-162.

[97] P.H. Winston, *Artificial Intelligence*, Addison-Wesley, Reading, MA, 1984.

[98] H.R. Alker Jr., J. Bennett, and D. Mefford, "Generalized Precedent Logics for Resolving Insecurity Dilemmas," *International Interactions* 7, 165-206 (1980).

[99] J.G. Carbonell, "Metaphor: An Inescapable Phenomenon in Natural Language Comprehension," in [119: 415-434].

[100] H.R. Maturana, "Biology of Knowledge," in R.W. Reiber, ed, *The Neurophysiology of Language*, Plenum Press, New York, 1977.

[101] I. Lakatos, *Proofs and Refutations* Cambridge University Press, Cambridge, 1976.

[102] M.A.K. Halliday, *Language as Social Semiotic*, Edward Arnold, London, 1978.

[103] H.R. Maturana, "Biology of Cognition," in H.R. Maturana and F. Varela, *Autopoeisis and Cognition: The Realization of the Living* Reidel, Dordrecht, 1980, 2-62.

[104] P.H. Winston, "Learning New Principles from Precedents and Exercises," *Artificial Intelligence* 19, (1982).

[105] J.J. Katz and J.A. Fodor, "The Structure of a Semantic Theory," *Language*, 39 (2), 170-210, (1963).

[106] R.C. Schank, "Conceptual Dependency: A Theory of Natural Language," *Cognitive Psychology*, 3, 552-631 (1972).

[107] E.A. Feigenbaum, and H.A. Simon, "EPAM-Like Models of Recognition and Learning," *Cognitive Science*, 8, 1984, 305-336.

[108] G. Duffy and J.C. Mallery, Relatus: An Artificial Intelligence Tool for Natural Language Modeling," AI Memo No. 847, Artificial Intelligence Laboratory, MIT, Cambridge, 1986.

[109] J.C. Mallery, "Constraint-Interpreting Reference," AI Memo No. 827, Artificial

Intelligence Laboratory, MIT, Cambridge, 1986.

[110] E.A. Feigenbaum, *An Information Processing Theory of Verbal Learning*, RAND, Santa Monica, CA, 1959.

[111] J.L. Kolodner, "Reconstructive Memory: A Computer Model," *Cognitive Science*, 7, 280-328 (1983).

[112] J.L. Kolodner, "Maintaining Organization in a Dynamic Long-Term Memory," *Cognitive Science*, 7, 243-280, (1983).

[113] P. Edwards, ed, *The Encyclopedia of Philosophy*, MacMillan, New York, 1967.

[114] H. Gadamer, *Philosophical Hermeneutics*,, D.E. Linge, ed & trans, University of California Press, Berkeley, 1976.

[115] F.R. Dallmayr and T.A. McCarthy, *Understanding and Social Inquiry*, University of Notre Dame Press, Notre Dame, 1977.

[116] H. Gadamer, *Hegel's Dialectic: Five Hermeneutical Studies*, P.C. Smith, trans, Yale University Press, New Haven, 1976. German edition published in 1971.

[117] J.B. Thompson and D. Held, eds, *Habermas: Critical Debates*, MIT Press, Cambridge, 1982.

[118] J.B. Thompson, ed and trans, *Paul Ricoeur: Hermeneutics and the Human Sciences*, Cambridge University Press, Cambridge, 1981.

[119] W.C. Lehnert and M.H. Ringle, *Strategies for Natural Language Processing*, Lawrence Erlbaum Associates, Hillsdale, NJ, 1982.