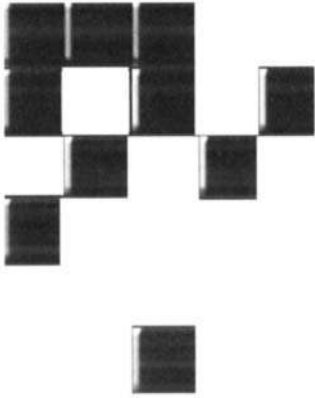# Language and Representation

Information Retrieval and the Philosophy of Language

David C. Blair
University of Michigan

Introduction

Information retrieval—the retrieval, primarily, of documents or textual material—is fundamentally a linguistic process. At the very least we must describe what we want and match that description with descriptions of the information that is available to us. Furthermore, when we describe what we want, we must *mean* something by that description. This is a deceptively simple act, but such linguistic events have been the grist for philosophical analysis since Aristotle. Although there are complexities involved in referring to authors, document types, or other categories of information retrieval context, here I wish to focus on one of the most problematic activities in information retrieval: the description of the intellectual content of information items. And even though I take information retrieval to involve the description and retrieval of written text, what I say here is applicable to any information item whose intellectual content can be described for retrieval—books, documents, images, audio clips, video clips, scientific specimens, engineering schematics, and so forth. For convenience, though, I will refer only to the description and retrieval of documents.

The description of intellectual content can go wrong in many obvious ways. We may describe what we want incorrectly; we may describe it

3

correctly but in such general terms that its description is useless for retrieval; or we may describe what we want correctly, but misinterpret the descriptions of available information, and thereby match our description of what we want incorrectly. From a linguistic point of view, we can be misunderstood in the process of retrieval in many ways. Because the philosophy of language deals specifically with how we are understood and *mis*-understood, it should have some use for understanding the process of description in information retrieval.

First, however, let us examine more closely the kinds of misunderstandings that can occur in information retrieval. We use language in searching for information in two principal ways. We use it to describe what we want and to discriminate what we want from other information that is available to us but that we do not want. Description and discrimination together articulate the goals of the information search process; they also delineate the two principal ways in which language can *fail* us in this process. Van Rijsbergen (1979) was the first to make this distinction, calling them "representation" and "discrimination."

The Retrieval Problem: Failures of Description

A failure of description can occur in a number of ways. The most obvious failure is when an item of information is described incorrectly: a textbook on "economics" is described, for example, as being on "anthropology," or a book by Mark Twain is described as being written by Henry James. But there are more subtle failures of description, too, such as when the description is generally correct but is beyond the comprehension of the typical inquirer who might see it. An example of this is a book described as being about "plate tectonics" when the typical inquirer who is interested in theories of "continental drift" may not realize that "plate tectonics" is the more formal description of the same subject matter. In other situations, opposing views arise as to how a particular literature should be described; for example, some researchers may consider "cold fusion" to be a valid field of scientific research deserving its own category, while others see work on "cold fusion" as more appropriately subsumed under the rubric "crank theories" or "pseudo-science."

When we look at just the reasonably correct or useful descriptions that can represent an item of information, this set of reasonable descriptions may be quite large. It has been shown empirically (Swanson, 1996), and argued theoretically (Blair, 1990), that the number of different descriptions that can represent the intellectual content of even a relatively short document may have no upper bound. This conclusion calls into question the notion of "exhaustive indexing"—the assignment of *all* the index descriptions that could represent the intellectual content of an item of information. Some have argued that information retrieval systems should use *all* possible index terms to represent the intellectual content of a document—a strategy called "unlimited aliasing" (Furnas, Landauer, Gomez, & Dumais, 1987). Such a strategy ignores two things. First, there may be no upper bound to the number of words and phrases that can represent the intellectual content of even a small item of information. Second, some of the many possible index terms will always be more useful for retrieval than others, so the assignment of *any* reasonable index terms to a document may not be the best indexing strategy—some index terms really are better than others, as Brooks (1993) has shown.

The high number of reasonable descriptions is both good and bad. It is good in the sense that it is easy to come up with one or more reasonable index terms. But it is bad in the sense that because so many reasonable descriptions for a document exist, a searcher may have a difficult time anticipating the ones actually assigned to the documents of interest, and further, documents that have the same intellectual content might be described in a number of different ways (e.g., one described as concerning "continental drift" whereas another on the same topic is described as concerning "tectonic plates").

The Retrieval Problem: Failures of Discrimination

Although the process of description is primarily focused on an individual document or category of information, the process of discrimination takes a broader view of the representation problem. It is not concerned only with individual documents or categories of information, but also with the relationship between the desired document(s) and the other documents that are available to the inquirer. The goal of discrimination is to

distinguish, by means of description, documents that are likely to be useful to the inquirer from available documents with similar intellectual content that are not likely to be useful. The ability to discriminate between useful and useless information establishes a continuum of description that can be characterized as ranging from specific (highly discriminating) to general (less discriminating) terms. The most obvious failure of discrimination is a description of the intellectual content of the desired document that is too general to distinguish it from the intellectual content of useless documents. For example, if the subject description "computers" were added to all the books and journals in a computer science library, it would have no discriminating power at all within that library. Such failure of discrimination is too obvious to be commonplace, but a more insidious form of discrimination failure can occur with even the most thoughtfully applied indexing descriptions. This failure happens when a description identifies a relatively small number of documents in an information retrieval system, and thus discriminates well, but during the lifetime of the system, more and more documents described in the same way are added. Eventually, the number of documents described in this way reaches a point at which the description, by itself, does not discriminate well enough to be of use to inquirers; that is, when the description is used by itself as a search term, it retrieves more documents than inquirers are willing to look through to find what they want (Blair, 1980).

Of course, the point at which a description fails to discriminate well is not a precise number and can be contingent on many factors, including the persistence of the inquirers using that description and the availability of other descriptions that can reduce the size of the less-discriminating category of information. Some inquirers are significantly more persistent or motivated than others and more willing to browse through large sets of retrieved documents. Such persistence often depends on the importance of finding the desired documents and the time available for the search. On the other hand, using other descriptions may reduce the number of items in a particular category. Time periods are frequently used to qualify a less-discriminating description, such as when one asks for only the most recent items described as being in the broad category of "computer science."

Recall and Precision

Any discussion of failure in information retrieval calls to mind the two complementary measures of retrieval performance: recall, which is the percentage of relevant documents retrieved, and precision, which is the percentage of retrieved documents that are relevant (Blair & Maron, 1985). In general, we can say that failures of description lead to low recall, whereas failures of discrimination tend to lower precision. Recall and precision are known to trade off in a rough and imprecise way— higher levels of recall are achieved at the expense of lower levels of precision, and vice versa. This effect suggests that description and discrimination may trade off in similar ways. Describing what we want in the most inclusive (that is, general) terms may lead to the construction of search queries that will be inclusive, but do not discriminate well (i.e., recall will be high and precision low). On the other hand, making our search queries as discriminating (that is, precise) as possible may lead to queries that do not describe what we want very inclusively (i.e., precision will be high and recall low). As in the case with recall and precision, the trade-off between description and discrimination is rough and imprecise.

The Processes of Description and Discrimination

The proliferation of electronic document collections, in particular the ubiquity of the World Wide Web (WWW, or Web), and the wide availability of Internet search engines have placed the tools of information retrieval in the hands of anyone with access to the Web. Individuals who in the past would have had to consult professional searchers such as librarians can now conduct their own searches. Such wide accessibility to public domain information can only please advocates of a free and open democratic society, but the widespread use of Internet search engines may be changing the way we ask for information. When inquirers asked a professional searcher for help in finding information, they could describe what they wanted using all the subtleties and nuances of natural language expression. The professional searcher, in turn, could clarify the inquirers' requests by asking appropriate questions. Now that typical inquirers conduct their own searches using search engines, much

of the subtlety of the interaction between inquirers and professional searchers has been lost. The typical information request submitted to an Internet search engine today is comprised of only a few words—often only one. As a consequence, it is important that we examine exactly what is meant by individual words when they are used to request (that is, to describe) information with a particular intellectual content.

Another change taking place in the information retrieval process is the dramatic growth in size of available document collections. Everyone is aware of how the Web is growing, of course, but even private document collections such as institutional and corporate intranets and document databases continue to grow at a spectacular rate. The reason for this is largely economic. We have reached the point with electronic document collections at which the cost of examining and discarding materials, such as Web pages that have outlived their usefulness, is higher than the cost of simply keeping them. As a result, we have many electronic collections that are never or rarely weeded of obsolete documents. The resulting collections of electronic information are, like the Internet, growing without any clear upper bound. But as document collections grow larger and larger, a subtle change in the information retrieval process is taking place. Instead of the goal of search query formulation being primarily the *description* of what is wanted, the overriding goal of query formulation has become the *discrimination* of small numbers of desirable documents from increasingly large numbers of unwanted documents.

What Do Descriptions Mean?

Because of the dramatic and seemingly inevitable growth in the size of information retrieval systems and the many ways that descriptions of information can go wrong, if we are to improve the complementary processes of describing what we want and describing what is available to us, it is important that we examine as closely as possible the activity of describing the intellectual content of information. At the very beginning of this discussion, I stated the obvious when I said, "when we describe what we want, we *must* mean something by that description." But what exactly *do* we mean when we describe what we want? A decade and a half ago, van Rijsbergen (1986a) wrote that one of the most conspicuously absent components of information retrieval theory had been

an explicit, formal notion of meaning. It is here that the philosophy of language may provide us with some guidance. (The complementary nature of the indexing and searching processes was a major theme of Blair [1990]. An earlier attempt to reduce the indeterminacy of these two processes was presented in Blair [1986].)

"Words and Meanings"

Philosophers have pondered the "meaning of meaning" since at least the time of Aristotle, but perhaps no philosopher has had more impact on the philosophy of language than Ludwig Wittgenstein (1889–1951). Wittgenstein's later work was instrumental in bringing about the "linguistic turn" in analytic philosophy during the 20th century. The "linguistic turn" resulted from the realization that philosophers who purported to study "ideas" were actually studying *descriptions of ideas*— not what we are thinking, but what we *say* we are thinking. The only direct access to ideas that we have is to our own ideas, by introspection. But we cannot easily generalize from our own introspection to statements about how others think (Hacker 1996b; Rorty, 1967). Wittgenstein (1953) reinforced this change in his later work, *Philosophical Investigations,* by arguing that many of the philosophical problems that puzzled philosophers were not problems at all, but were merely the result of misuses of language. As he put it so succinctly, "Philosophy is a battle against the bewitchment of our intelligence by means of language" (Wittgenstein, 1953, p. 47). It would be impossible, of course, to provide a complete discussion of Wittgenstein's extensive work in the philosophy of language here. His published works run to 13 volumes, and his *Nachlass,* or literary estate, much of which is still not published, is even larger—more than 30,000 pages (this is in the process of being published in 15 or more volumes as the *Wiener Ausgabe* [Nedo, 1993]). The story of the complications and intrigues of this project are detailed in Toyton (1997). The complete electronic CD-ROM versions of Wittgenstein's *Nachlass,* published writings, lectures, and letters, are each available from InteLex (http://library.nlx.com/). The reader should also understand that Wittgenstein is not the only major philosopher of language; I will discuss some others here. Nor is it the case that every philosopher accepts Wittgenstein's arguments without dispute. My purpose here is not to defend Wittgenstein, but to

present relevant portions of his work as clearly as possible because it has been enormously influential in philosophical circles and in related areas of linguistics and psychology. An excellent overview of 20th century philosophy of language is provided by Lycan (2000). Blackburn (1984) provides an introduction to the philosophy of language written specifically for the nonphilosopher. Devitt and Sterelny's (1999) introduction to the philosophy of language includes a section on "language and mind" and a discussion of linguist Noam Chomsky's work. Finally, many of the salient papers in the philosophy of language are collected in Rosenberg and Travis (1971). A more recent collection can be found in Ludlow (1997).

The commentaries on Wittgenstein's work are also extensive. The most detailed commentaries on Wittgenstein's most influential work, *Philosophical Investigations,* are by the co-authors G. P. Baker and P. M. S. Hacker (the first two volumes are by Baker and Hacker [1980, 1985], and volumes 3 and 4 are by Hacker alone [1990, 1996b]). Wittgenstein's discussions on specific topics are frequently scattered throughout his writing, so Garth Hallet's (1977) concordance to *Philosophical Investigations* can often be an extremely useful tool for locating and bringing together his writings on the same subject. Wittgenstein's former student and Cornell philosophy professor, Norman Malcolm, has put together several collections of his own insightful essays on Wittgenstein's work. Of particular note is his *Wittgensteinian Themes: Essays 1978–1989* (Malcolm, 1995). Even Wittgenstein's personal life has a compelling interest because, for Wittgenstein, philosophy was not just a collection of puzzles, but a guide for living; as he once said,

> what is the use of studying philosophy if all that it does for you is to enable you to talk with some plausibility about some abstruse questions of logic, etc., if it does not improve your thinking about the important questions of everyday life? (Malcolm, 1972, p. 39)

The two best biographical works are Norman Malcolm's (1972) short but intimate *Ludwig Wittgenstein: A Memoir,* and Ray Monk's (1990) excellent, detailed biography *Ludwig Wittgenstein: The Duty of Genius.* A third work, Theodore Redpath's (1990) *Ludwig Wittgenstein: A*

Student's Memoir offers an undergraduate's impressions of the philosopher. Several of Wittgenstein's students have published literal transcriptions of his classroom lectures and discussions (Ambrose & Macdonald, 1979; Geach, Shah, & Jackson, 1989; King & Lee, 1978). Finally, Bouwsma published his notes of discussions that he had with Wittgenstein during the last few years of the philosopher's life (Bouwsma, Wittgenstein, Craft, & Hustwit, 1986).

An important characteristic of Wittgenstein's work was his own self-criticism. Early in his career he was strongly influenced by the logic and analytical philosophy of Gottlob Frege and Bertrand Russell. After studying with Russell at Cambridge University, he wrote his first book, *Tractatus Logico-Philosophicus* (Wittgenstein, 1961a, 1961b). This was the only philosophy book by Wittgenstein published during his lifetime; it lays out a rigorous, logical model of language and a "picture theory of meaning" that has clear antecedents in the work of Frege and Russell. Wittgenstein wrote most of the *Tractatus* while serving as a much-decorated soldier in the Austrian army during World War I. After writing the *Tractatus,* Wittgenstein felt that he had solved the major problems of analytic philosophy. But while he was away from academic life, his book was having a major impact on analytic philosophy in England, of course, but also within the newly formed Austrian movement in analytic philosophy, Moritz Schlick's "Der Wiener Kreis" (The Vienna Circle). Wittgenstein began to see that he had not solved all the problems of philosophy and that there were serious problems with some things he said in the *Tractatus*. He spent the remainder of his academic life at Cambridge University. Although Bertrand Russell strongly supported his return to professional philosophy, Wittgenstein was soon to criticize and change much of his earlier philosophy that Russell had found so attractive. Wittgenstein's (1953) reassessment of his early philosophy culminated in the collection of philosophical remarks called *Philosophical Investigations*. Although the *Investigations* was a product of Wittgenstein's extensive editorial efforts over the last years of his life, it was not published until shortly after his death, and the questions it raised, of course, could not be answered by Wittgenstein himself. (Wittgenstein's other books have been put together from selections of his unpublished writings by his former students after his death.) Wittgenstein has left us with two interpretations of his intellectual

legacy: *Philosophical Investigations* is either an extensive critique of his earlier work in the *Tractatus,* or, as some commentators insist, the second of two largely separate philosophies. Which of these two views is correct will probably never be answered to everyone's satisfaction, but the best attempts to put Wittgenstein's early and late philosophy into perspective are Norman Malcolm's (1986) *Nothing Is Hidden: Wittgenstein's Criticism of His Early Thought,* and P. M. S. Hacker's (1989) *Insight and Illusion.*

Although Wittgenstein focused his philosophical efforts on many specific themes, the published works themselves do not separate his writings into categories: *Remarks on the Foundations of Mathematics* (1978) contains many remarks on language as well as mathematics, *Philosophical Investigations* (1953) contains remarks on philosophy in addition to mathematics, logic, and psychology, and *Remarks on the Philosophy of Psychology* (1980) contains remarks on language, psychology, and other subjects. Wittgenstein's philosophy on a particular topic is, in some sense, everywhere in his writings, but not collected or summarized in any one place. Successive paragraphs in a given work may deal with a specific topic, but the topic is dropped in favor of another and picked up again, seemingly at random, later in the work, or in another work. Certainly one of the reasons for this patchwork approach to philosophy is that Wittgenstein was continually grappling with very deep and elusive problems, problems that had defied systematic solution by the best analytical minds of the 20th century. So, many of his recorded comments were not solutions to these problems, but the remnants of an intellectual battle that he fought all his life (his published writings go right up to a few days before he died, when he succumbed to a long illness). Those who are interested in specific aspects of Wittgenstein's work and do not have the time to make a study of his extensive writings must rely on secondary sources to bring together his work on particular topics. Fortunately, there are some good works. For those interested in his late philosophy of language, the first 130 pages of Hanna Pitkin's (1972) *Wittgenstein and Justice* is, in this author's opinion, the single best introduction to this aspect of his work. Those interested in Wittgenstein's thought concerning more specific topics such as the determinacy of sense, the rejection of private languages, the denial of psycho-physical parallelism, and the rejection of mind–body dualism, among others, would do well to consult Glock's (1996) *A Wittgenstein Dictionary,* which

contains short discussions on, and references for, many major and minor topics in Wittgenstein's writings. Readers who would like to see Wittgenstein's writings on the same topic, but in different works, brought together should consult Anthony Kenny's (1994) *The Wittgenstein Reader.*

Wittgenstein's writings on the philosophy of language were extensive and closely linked to his views on the philosophy of mind. To him, language is not a *product* of thought, as most philosophers accepted; "language," as he put it, "is ... the vehicle of thought."

> When I think in language, there aren't "meanings" going through my mind in addition to the verbal expressions: the language is itself the vehicle of thought. (Wittgenstein, 1953, p. 107)

Or, stated somewhat differently:

> Knowledge is not *translated* into words when it is expressed. The words are not a translation of something else that was there before they were. (Wittgenstein, 1967, p. 32)

The point Wittgenstein is making is not that all thought uses language as a medium, for we can surely "think about" music or visual images without reference to language at all, but that when we use language we usually use it as a means for thinking, not as a product of thought or as an expression of something we "have in mind."

This "Copernican Reversal" in the way that thought and language had traditionally been seen to be related has important implications for information retrieval. The process of information retrieval is often seen as one in which the inquirer has something "in mind"—an "information need"—which he or she then translates into an actual search query, in the same way that people were thought to express in ordinary language what they already had "in mind." But if Wittgenstein is right that our use of language is a form of thinking, then the "language" of retrieval— the search terms that are available to us and the ways in which they can be combined—are the "language" with which we think about, and thereby articulate, what information we want. In short, how we think about our information needs is strongly constrained by the retrieval

language that is available to us, and insofar as the language of retrieval is limited, so will be our thinking about what we want. The language of retrieval not only limits how we articulate what we want but can also constrain the very thought process in which we determine what we want. Presumably, we would like to think that we mold our information retrieval systems to serve our need for finding information; but, if Wittgenstein is correct, then it may be the case that our information retrieval systems are molding us to think along their lines. If this is the case, then it may be extremely difficult to design radically different or improved retrieval systems, because we are virtually locked into the way of thinking about retrieval that is embodied by existing systems.

Wittgenstein presented his own view of language in terms of a critique of traditional, widely accepted views. In *Philosophical Investigations* he presents a theory of language based on the writings of the medieval philosopher St. Augustine. The Augustinian model of language is a simple referential model that, although old, has been remarkably persistent, existing in various forms even today. The Augustinian model of language sees linguistic meaning in the following way:

1. Words name objects: the meaning of a word is the object for which it stands.
2. Every word has a meaning.
3. The meaning of a word is independent of context.
4. Sentence meaning is composed of word meanings.

1. Words Name Objects

If we consider examples of words such as "chair," "apple," and "pencil," language does seem to work this way. But if we look at examples such as "rectitude," "charisma," and "the day after tomorrow," it is harder to make the case that words name "objects." Wittgenstein (1953, p. 174) gives us a hint of the complexity he sees in these kinds of statements when he asks rhetorically:

A dog believes his master is at the door. But can he also believe his master will come the day after to-morrow?[sic]—

> And *what* can he not do here?—How do I do it?—How am I
> supposed to answer this?

If "the day after tomorrow" were simply a phrase correlated with an object, or "meaning," of some kind, it would be plausible that even a dog could understand it and could come to expect his master then. A dog, after all, can recognize other kinds of objects: balls, bones, food, leashes, cats, and other dogs, as well as more abstract objects such as friends and enemies, and characteristic situations like his master coming home soon, playing, or being frightened. Further, a dog can "expect" things to happen some short time in the future, such as getting fed. But can a dog expect his master "the day after tomorrow?" Wittgenstein does not answer his question explicitly, but it is clear from his writings that he does not believe that a dog can do this. For Wittgenstein, the "day after tomorrow" is not a phrase that has a meaning, although we sometimes speak of it this way; that is, there are circumstances in which we use the word "meaning" in this way (we can imagine someone who is learning English asking "What does the 'day-after-tomorrow' mean?"). According to Wittgenstein (1953, p. 20), if we really want to understand the "meaning" of the "day after tomorrow" we need to look at its use:

> For a large class of cases—though not for all—in which we
> employ the word 'meaning' it can be explained thus: the
> meaning of a word is its use in the language.

Consequently, to understand the "meaning" of the "day after tomorrow," we need to be able to use it in the right circumstances, and to use it in the right circumstances we need experiences of distinguishing one day from another—"today," "tomorrow," "the day after tomorrow"—of observing the succession of one day following another, and of using days as units of time in a variety of activities. Further, these are not independent activities that can be separated from our daily lives and practices. To use the "day after tomorrow" correctly is not just to know a dictionary definition, it is to be able to discern the appropriate circumstances and activities in which it can be used, and this ability is further contingent on our ability to participate in a broad range of human activities in which understanding the "day after tomorrow" is important.

Someone who speaks another language and is learning English asks me, "What does the 'day after tomorrow' mean?" He can get along with this simple question because he already speaks another language and is probably familiar with the kinds of activities in which such a phrase is used. A dog, however, doesn't share with us the activities in which the "day after tomorrow" is important. Wittgenstein (1953, p. 223) brings this out more strikingly with one of his more enigmatic statements:

> If a lion could talk we could not understand him.

The reason we could not understand the speaking lion is that we have no personal experience of the activities in which he is engaged. If we can come to understand the meaning of a word by looking at its use, then meaning is intimately linked to the activities and practices that we have in common with others. If we do not have any activities in common, then there is nothing that we can talk about. In Wittgenstein's words, we have too few "forms of life" in common with the lion on which we could base a common language. For Wittgenstein:

> We don't start from certain words, but from certain occasions or activities. (Wittgenstein, 1972, p. 3)

> Only in the stream of thought and life do words have meaning. (Wittgenstein, 1967, p. 30)

After considering these problems with the Augustinian model of language, we may try to draw some comfort from the fact that language appears to work according to Augustine's model, at least in the cases where actual objects are referred to. But even here the relationship between language and "objects" is not simple. "Words" and "objects" recall the much-debated topic of "reference." Frege was one of the first philosophers to discuss some of the complexities of reference, but these issues reach back at least to the third century B.C. and Eubulides' "paradox of the masked man" (sometimes called "the paradox of the hood"). Suppose, said Eubulides, that you see a masked man. In reality, the masked man is your brother. But you cannot say that you saw your brother. Frege (1952) highlighted one of the important issues of reference with his example of the "Morning Star" and the "Evening Star."

Both the Morning Star and the Evening Star refer to the same celestial object, the planet Venus. Yet the descriptions "Morning Star" and "Evening Star" do not have precisely the same meaning for the simple reason that we cannot use them interchangeably in everyday discourse. That is, in ordinary usage we cannot say, in the morning, that we see the "Evening Star" and, in the evening, that we are looking at the "Morning Star," although neither statement is, technically, false.

In Frege's example we can at least tell what the speaker who refers to the "Morning Star" in the evening, or the "Evening Star" in the morning actually *means*. But Bertrand Russell (1905, p. 485) gave us an example of a problem of reference where it is not at all clear what the speaker means. Consider the following two statements:

"George the IV wished to know if Scott was the author of *Waverley*."

"Scott is the author of *Waverley*."

Now if "Scott" and "the author of *Waverley*" refer to the same person, *and* the meaning of a word is completely explained by its reference, as Augustine claims, then we should be able to use "Scott" and "the author of *Waverley*" interchangeably. If we substitute "Scott" for "the author of *Waverley*" in Russell's first sentence, then we get: "George the IV wished to know if Scott was Scott." Here, in contrast to Frege's example, the substitution of "Scott" for "the author of *Waverley*" leaves us with a sentence that Russell believed was clearly false and whose intended meaning would be impossible to discern. Russell (1905) went on to develop his *Theory of Definite Descriptions*, which was aimed at uncovering the logical form (as opposed to the grammatical form) of statements that refer to a single individual, like "the author of *Waverley*." This was used to address some of the troublesome puzzles about definite descriptions, such as the substitutivity problem, above, and references to nonexistent things (e.g., "There is no place called Shangri-La"). Although such a level of detail is beyond the scope of this review, it still makes engaging reading. (The interested reader should consult chapter 2 of Lycan [2000] for a very readable presentation of Russell's work on definite descriptions

and the subsequent debates that he engaged in, principally with Strawson [1950].)

It is clear that even when a word or phrase has an obvious reference such as "Scott" and "the author of *Waverley*" do, the sense, or "meaning" of that word or phrase is more than just its reference. In some cases when we refer to a particular person we may not mean the person at all, but some salient aspect of the person. For example, Wittgenstein's father, Karl, was once referred to as "the Andrew Carnegie of Austria." By this, it was *not* meant that Karl looked like Carnegie, or had Scottish ancestry, but that he, like Carnegie, was a wealthy industrialist who patronized the arts. Finally, it is evident that many words, such as "rectitude" and "unicorn" do not refer to objects at all, yet we still use them regularly and are understood when we do so. Meaning must be something other than simple reference.

Augustine's model of language is a simple model and easy to comprehend, but some subtle aspects of it are not obvious at first. In particular, Augustine's description of how he learned to speak is important. In his words, he "...heard words repeatedly used ... [and] gradually learnt to understand what objects they signified ..." (Wittgenstein, 1953, p. 2). This passage makes the point that we can hear and distinguish words before we understand them. That is, words can exist for us without meaning—as words we don't understand. Further, because we can have words without meaning, it follows that "meaning" can exist independently of words—it appears to be something that can be added to words by a specific act such as looking them up in a dictionary. In some instances we can even have a "sense" or "meaning" without a word. We can see this sometimes when we compare words in two languages. For example, the Japanese have a word that means the point when a sound, such as the single stroke of a large bell, has diminished to a level where the listener cannot tell whether he can still hear it or not. In English, we don't have a word or simple phrase for this "meaning." Augustine's view of language dichotomizes words and meaning and sets up a framework in which they can be considered separately, a framework that exists in various forms to this day, most prominently in the belief in the independence of syntax and meaning that was the cornerstone of Chomsky's (1965) generative grammar.

This dichotomy between words and meaning forces us to deal with questions of meaning in a predictable, almost unavoidable, way. Specifically, when we can no longer maintain the claim that meaning equals some entity such as an "object," we give up the "object" but we inevitably try to keep the framework in which the "meaning" of a word is an entity of *some* kind. We think of a word having a "meaning" in the same way that we think of people having biological parents. The child may not know who his parents are, but their existence at some time is beyond doubt. Wittgenstein, too, believed in the dichotomization of meaning and words or grammar early in his career. But it was one of his major contributions to the philosophy of language to question this fundamental dichotomy; in short, to resist the "compulsion" to separate words and meaning that the Augustinian model of language seems to force on us.

> The questions "What is length?," "What is meaning?," "What is the number one?" etc., produce in us a mental cramp. We feel that we can't point to anything in reply to them and yet ought to point to something. (We are up against one of the great sources of philosophical bewilderment: a substantive makes us look for a thing that corresponds to it.) (Wittgenstein, 1958, p. 1)

Augustine's model of language reinforced the basic dichotomy between words and meaning, leaving to subsequent philosophers the task of trying to get them back together again. The fact that many words and phrases obviously do *not* refer to objects, yet are nonetheless meaningful, compels us to look for another entity that "meaning" could be. John Locke was able to articulate an alternative theory of meaning that preserved Augustine's separation of words and meaning but did not fall prey to its failures. For Locke (1690/1985, p. 114):

> Words in their primary or immediate signification, stand for nothing but the ideas in the mind of him that uses them.

Locke's linking of a word's signification, or "meaning," with an "idea" resolves the problem of words that do not refer to a physical object or type of object. We may not be able to link all words to objects, but it

seems evident to some theorists that when we know the meaning of the words "rectitude" or "unicorn" we do have something "in mind." It is then easy to take the next step and assume that what we have "in mind" is what the word *means*. This is the "mentalistic theory of meaning"—a semantic theory that has widespread appeal and is implicit in much information retrieval theory.

Locke's mentalistic theory of meaning has had a long history of support, and various forms of it survive today. But, as appealing as mentalistic theories of meaning are, they suffer from a number of fatal problems. In the first place, if the meaning or sense of a word that I understand is an idea, then that idea, by definition, is something private to me. But if meaning is private, how do I teach you *my* idea of the meaning of a word, or learn the meaning of a word that you understand but I do not—after all, you cannot see what is in my mind. We *do* explain the meaning of words and phrases to each other, but is this explanation really a presentation of our ideas? If the explanation we give turns out to be wrong, what is the source of our error? Did we have the right idea, but explained it incorrectly, or was our original idea incorrect in the first place? There is no way to tell. Yet to teach or learn the meaning of a word or phrase requires clear criteria of correctness, something a purely mentalistic theory of meaning does not—and cannot—have. For Wittgenstein, the criterion for whether you understand the meaning of a word is not whether you have the right idea, but whether you use it correctly in your day-to-day speech and writing. If I want to teach you the "meaning" of a word, I can give you examples of how it is used, or show you how it is used in the appropriate actual or hypothetical circumstances. The question of whether you have the "right idea" doesn't come up in ordinary usage. Thus, if the criterion for correct understanding is correct usage, then ideas are not the foundation of our understanding—usage is. This is not to deny that some "mental phenomena" accompany our language use, it only means that, whatever those "mental phenomena" are, they are not required for teaching or learning a language; they are what epistemologists call "epiphenomena." The problem in semantics is not what the definition of "meaning" is, the problem is the seeming dichotomy between words and meanings that encourages us to think of "meaning" as a separate entity—something that can be linked to words and examined apart from usage. This is an example of what

Wittgenstein (1958, p. 143) called a "disease of thinking." A disease of thinking is a mistaken way of conceptualizing a problem that leads us unavoidably to the wrong conclusion. This is exactly what happens, Wittgenstein says, when we dichotomize our view of language by saying that "words have meanings." As soon as we talk as if there are such things as "meanings" that are linked, somehow, to words, we quite literally force ourselves to grant the independent existence of "meanings." Once we do that, it is a forgone conclusion that we will find something that we will be able to call the "meaning" of a word. Wittgenstein does not discuss the notion of a "disease of thinking" at any length, but the philosopher Gilbert Ryle (1931, p. 139), who admitted to being strongly influenced by Wittgenstein, wrote a classic paper in which he describes and discusses a number of different kinds of linguistic errors like this— what he aptly called "systematically misleading expressions." Note, though, that asking for the "meaning" of a particular word, is a quite ordinary and acceptable kind of request for speakers to make. Wittgenstein felt that it was acceptable for ordinary speakers to talk this way because they weren't concerned about the ultimate status of "meanings." The problem occurs when philosophers try to analyze this statement. For Wittgenstein, many of the "diseases of thinking" about language are only problems for philosophers, not for ordinary native speakers.

But "meanings" are not separate things that can be examined as a geologist examines rock samples. Meanings are not "entities," but rather are emergent phenomena arising from our day-to-day activities and practices (Holland, 1998). They are not solely mental entities, conscious or unconscious, because they are usually contingent on the circumstances and context of their usage. But although Wittgenstein linked meaning and use, he did not intend for meaning to be interpreted solely as behavior, a common misunderstanding of his work. Meaning is not solely behavioral because it often has a mental component, otherwise we would not be able to distinguish between someone who lies but has the same statements and behavior as someone telling the truth. Wittgenstein (1953, p. 220) expressed his attitude toward meaning and use most clearly when he said, "Let the use of words teach you their meaning."

The philosopher Hilary Putnam (1988) identifies several other problems with mentalism in his essay "Three Reasons Why Mentalism Can't

Be Right." Max Black (1968), a contemporary of Wittgenstein, presents an argument for the rejection of mentalism in his *The Labyrinth of Language*.

2. Every Word Has a Meaning

Even if the "meaning" of a word is not an object, there is still a tendency to think of the meaning of a word as a single thing, something that is the same in all applications. This is implicit in the notion that meaning is fixed by definitions, a view with which Wittgenstein explicitly disagrees. Again, if we think of tangible objects—chairs, cars, hammers, and the like—this view has a certain appeal. But on closer examination we can see that, even with common objects, there can be cases in which the definition or meaning is not a single thing. Those who think that a chair is simply a chair, should go to a museum of contemporary art. Here, what an artist may call a chair can vary widely from our accepted notion of what a "chair" is. But even in ordinary usage, what we might call a chair can deviate from our normal expectations. A chair has a function, it is something to sit on. Such a function can give the status of "chair" to a lot of objects. For example, if we need to sit down, but there are no ordinary chairs available, we might use a low table or a box to sit on. In a functional sense, the low table or box becomes a chair for the period of time we use it for this purpose. What is important is that the definition of even an ordinary object like a chair is not fixed. The boundary between what is a chair and what is not a chair may be unclear, and may vary according to circumstances. In a like manner, we can view a hammer as a specific kind of tool with a characteristic shape and heft, but we can also view a hammer as something that can be used in certain ways. In the functional sense, a lot of things can be used as hammers: rocks, iron bars, even fists.

The words "hammer" and "chair" can also be used metaphorically, or as figures of speech, such as when the weatherman says that a storm "hammered Cape Cod" or when a reporter states that Senator X "chaired the Armed Services Committee." These metaphorical or figurative uses of these words stretch our notion of what they mean and where it is appropriate to use them.

Some names of objects find a wide variety of applications. The word "head" is used to denote a particular anatomical feature of most animals,

but there are other related uses of "head" that may be only distantly related to this anatomical feature:

> "He went to the **head** of the line."
> "She's the **head** of the executive board."
> "The crisis quickly came to a **head**."
> "They began the canoe trip at the **head** of the river."
> "The outlaws wuz **headin'** North, sheriff!"
> "The sailor's punishment was to keep the **head** clean for the
> next two weeks."

Such examples do not exhaust the different uses of the word "head," and this variability is typical of many other common words, such as "line" or "pitch." Examples like these should dispel any notion that word meaning is precise or a single thing or can even be limited to the various definitions listed in an unabridged dictionary. Hilary Putnam (1988) goes even further. In the first place, he says, meaning is "holistic." By this he means that the meaning of a word is not fixed once and for all by a definition (as logical positivists insisted it had to be), but is contingent on how it is used in a wide variety of statements and circumstances. No single use is definitional. Further, the meaning remains "nonmonotonic" or "defeasible"—it is always subject to revision or change. For example, when I say that the word "bird" means, in part, feathered bipeds that can fly, I do not mean for my listeners to conclude that a bird with a broken wing can fly, or that a newly hatched bird can fly. Nor do I mean that birds like penguins or ostriches can *ever* fly. It is also the case that there may be future unanticipated circumstances in which the birds will not be able to fly. For example, we might find that if a bird were taken into space aboard the space shuttle, it could not fly in weightless conditions. It is also possible that genetic engineering will produce flightless birds of a new type.

Putnam (1988) goes on to show that even when we agree on the meaning, or usage, of a particular word we still may not all have the same criteria for its use. Meaning, he proposes, is subject to a kind of "division of linguistic labor." By Putnam's account, even when we agree on the "meaning of a word" we may not be using identical criteria for its application. I may recognize an elm tree by the general size and shape of its serrated

leaf, while my friend may recognize elms just as reliably as I do by examining the shape of the mature tree, the appearance of the bark, and the characteristics of the leaf buds. An expert botanist might be able to identify an elm by the particular cell structure of the wood, which can be seen under a microscope. Putnam's point is that the ability to use a word—here, "elm"—in the same way does not guarantee that the users possess the same criteria for the word's usage. Language, according to Putnam, is a cooperative activity. We may have useful heuristics that help us identify things like elms, but no one, not even the expert, can identify things like elms in every possible circumstance. For example, a botanist would probably not be able to identify an elm in complete darkness, but a blind person may have touch sensitive enough to do so by handling the bark and leaves. To distinguish an elm from a tree that looks very similar to it, or to identify an elm in the winter when it has no leaves, we would probably rely on expert botanists or a tree identification guide. But if we know very little about elms, we may just rely on our neighbor to help us identify them. This is what Putnam means by the "division of linguistic labor." The expert, according to Putnam, does not know a "more complete" definition of "elm" than we do, he or she simply knows more about elms than we do, and this additional information about elms may be useful for identifying elms in certain circumstances.

Wittgenstein (1953) would say that what accounts for the different criteria that we have, even when we can each identify elms reliably, is that we use the word "elm" in different "forms of life" and "language games." Forms of life are the regular activities and practices we engage in on a day-to-day basis, and language games are the regular patterns of word usage that dictate how we employ language in these activities and practices. Certain language games may be used primarily in specific forms of life. We need only the criteria to identify elms in the forms of life and language games that concern us. If we are not botanists, we may want only to identify the elms that we see in our own yard or neighborhood. The ability to distinguish species of elms, e.g., those that grow in southern latitudes from those that grow in northern latitudes, may not be important to us. For some botanists, distinguishing various species of elm may be important, but understanding the cell structure of elm wood may not be important. For a botanist at a tree farm, understanding the many varieties of cultivated elm species may be important, but understanding how

to identify elms in the wild may be unimportant. And a botany student studying for an exam may need only to be able to write down the scientific definition of an "elm," which might include its Latin name and its correct phylum and genus. Some of these language games may pick out the same trees as elms, and some may not. There is no sense of a "complete" definition of an "elm" that would enable us to pick out elms in all conceivable circumstances because a single individual would probably not find himself or herself in "all conceivable circumstances." It also should be clear that there is no single language game that would require such an all-inclusive understanding of what an elm is. The criterion for whether individuals know the meaning of a word like "elm" is not whether they command some arbitrarily "complete" definition of an elm, but whether they can use the word "elm" correctly in the activities and practices in which they wish to participate.

3. The Meaning of a Word Is Independent of Context

Indexicals are good examples of context-dependent words: words like "here," "now," "this," "that," "him," "her," and "it." The references for these words change from context to context. These examples are fairly obvious, but other examples are more subtle and deal with aspects of context beyond the notion of physical presence or absence. Wittgenstein (1969, p. 348) gives the example, "I am here." This sentence has the indexicals "I" and "here," and these would be clarified by ascertaining who spoke the sentence and on what occasion. But if, as Wittgenstein notes, you are sitting before me and are perfectly visible, and you utter, "I am here" you probably mean something else entirely than the simple statement describing where you are; that is, it is obvious to me that you are here, so you must be trying to tell me something else. One can imagine a situation in which one person is distraught over something. A close friend or relative approaches, touches his or her hand, and says, "Don't worry, I am here," meaning, of course, not just that the speaker is physically present, but that the speaker is emotionally supportive. In this utterance the context needed to interpret the meaning of the sentence extends beyond the simple notion of physical presence and

includes the relationship between the two individuals and the particular circumstances in which they find themselves.

Context can often indicate which of the many meanings of a word is currently being used. For example, the word "pitch" can mean a lot of things: the slope of a roof, the modulation of voice, a specific action in a baseball game, a tar-based substance, the description of a product that a salesman gives his customer, and so on. But if two individuals are talking while they attend a baseball game and use the word "pitch," it is highly unlikely that it means anything other than a specific action in the game. In fact, if the speaker at the baseball game were to continually remind the listener that he or she was using the word "pitch" to refer to an action in the baseball game and not any of the other uses of "pitch," such explanation would be considered bizarre, irritating, or even insulting.

Sometimes the context of an utterance can be so strong that it completely overrides the meaning of the actual words spoken. President Franklin Roosevelt would often dispel the boredom of a long receiving line before a White House dinner by saying completely inappropriate things to the guests as he greeted them—one of his favorite greetings was "I murdered my grandmother this morning." The guests, of course, would not hear his exact words and would assume that the President had greeted them in a cordial and expected manner (Fadiman & Bernard, 2000). Some individuals can even carry such a strong context with them that it becomes virtually impossible for them to say, or be understood to say, certain things. Could the Dalai Lama make an obscene gesture? Probably not, no matter what his intention actually was.

4. The Meaning of a Sentence Is Composed of the Meanings of Its Words

If we insist on understanding "meanings" as entities that are somehow attached to words, it can lead us into another problem; specifically, we may conclude that sentence meaning is somehow put together from the meanings of the words. If this is the case, then, most fundamentally, understanding the meaning of a sentence means that we must *be aware* of all the words in a sentence. Yet there is ample empirical evidence that, at least with speech, we often do not even hear every word in a sentence.

For example, suppose that you walk into a fast food restaurant and go up to the counter to order what you want. As you approach the counter, a clerk approaches you and says, "Kelp ya?" In spite of the fact that the clerk, strictly speaking, has not said any English words, we generally understand him to have said, "Can I help you?"—a phrase that makes sense. In cases like this, it is clear that we understand the situation or circumstances *before* we understand what is said to us, and our expectations about what is likely to be said may supersede what is actually said. Thus, sentence meaning, in this example, is not built up out of word meanings for the simple reason that we did not hear the words in the sentence.

But what about cases where we *do* hear all the words in a sentence. Are there ever cases in which the meaning of a sentence seems to have nothing to do with the meanings of the individual words? Yes, there are. Consider the following example. I come into my office in the morning, and after greeting a colleague I ask her, "Is Bill back from vacation yet?" She answers, in a perfectly forthright manner, "I saw a yellow Volkswagen in the parking lot this morning." What did she mean? It is not difficult to determine that she is saying that she believes Bill is back from vacation—i.e., that Bill owns a yellow VW, that a yellow VW is uncommon enough that Bill probably has the only one, and if Bill is back from vacation he will have likely driven his car to the office and parked it where it can be seen by others. But no matter how extended and detailed the descriptions are of the meanings of "saw," "yellow," "Volkswagen," "parking lot," and so on, there is no way that we can derive the meaning my colleague intended only from the meanings of her words. To understand what is happening here, we must turn to the work of another philosopher of language, Paul Grice, who distinguished between sentence meaning and speaker meaning.

The idea that words have meanings or definitions and that we come to understand what is said by somehow "looking up" the meanings of what we hear has been called the "Coding Theory of Language" (Eco, 1976). In short, words are codes that we must read or hear and then translate in order to arrive at their meaning. Sometimes language works this way, but, according to Grice, it is often the case that our understanding of what is said or written is *inferential*. Specifically, when we listen to what is said to us we begin with a number of assumptions about

the intentions of the person to whom we are listening. Grice called these assumptions "conversational implicatures." In this example, when our colleague tells us "I saw a yellow Volkswagen in the parking lot this morning," she does not answer our question directly, and we must therefore infer what she meant rather than take the literal meaning of what she said to be what she intended. But we can only do this if we make the initial assumption that she is honestly trying to answer our question. This is the "cooperation principle." Whenever we talk to someone, we almost always assume they want to cooperate with us. But we make other assumptions too. Grice's (1989, pp. 26–27) Principle of Cooperation is based on the satisfaction of nine maxims, which fall into four categories: Quantity, Quality, Relation, and Manner:

> There are two maxims of **Quantity:**
> 1. Make your contribution as informative as is required (for the current purposes of the exchange).
> 2. Do not make your contribution more informative than is required.
>
> Two maxims of **Quality:**
> 1. Do not say what you believe to be false.
> 2. Do not say that for which you lack adequate evidence.
>
> One maxim of **Relation:**
> 1. Say only what you believe to be relevant.
>
> Four maxims of **Manner:**
> 1. Avoid obscurity of expression.
> 2. Avoid ambiguity.
> 3. Be brief (avoid unnecessary prolixity).
> 4. Be orderly.

It is my assumption, in this example, that my colleague is answering my question in a cooperative and relevant manner that allows me to make the necessary inferences that lead me to understand what she means. Without this assumption of cooperation, our communication would be far more difficult, and everything we say would have to

be taken literally; that is, statements whose literal interpretation did not make sense would have to be considered odd, inappropriate, or meaningless.

In information retrieval, the process of describing what we want and evaluating what we retrieve is a lot like a conversation. We make requests and the retrieval system "answers" with sets of retrieved documents. Thus, we would expect that for successful "communication" to take place, Grice's maxims must be upheld in the search process. It is here that we can get the clearest picture of the difference between using a search engine to find information and asking an experienced searcher, such as a librarian, to help us. For both a search using a search engine and a search in which we ask a professional searcher to help us, we will probably assume, as Grice asserts, that both will be cooperative—they will try to answer our request. The difference between the two situations becomes clearer when the initial search fails to produce useful information. With the professional searcher, we can explain which of Grice's maxims has been violated in the search, and thereby provide guidance to the searcher on how to revise the search. For example, if we don't get enough information (a violation of the first maxim of Quantity) we might say, "That's what I'm looking for, but I need more of the same kind of information." If we get too much information (a violation of the second maxim of Quantity), we might say, "That's way too much detail, can you get me a more concise summary?" In another situation, we might receive information that is on the desired topic, but of questionable veracity (a violation of the first maxim of Quality), so we might say, "The information you got me claims that event X actually occurred, but can you find any documentation that substantiates that claim?" We could give examples for the violation of Grice's other maxims, too. The only kinds of retrieval failure that have been discussed much in the information retrieval literature have been the retrieval of nonrelevant documents (a violation of the maxim of Relation) and the retrieval of too much information (a violation of the second maxim of Quantity). It might be useful to consider the other maxims as additional criteria for successful retrieval (Blair, 1992).

Externalism and the Philosophy of Language

Recent work in the philosophy of language has shown the influence of the trend toward "Externalism" in the philosophy of mind. (A good presentation of the various forms of Externalism can be found in McGinn, 1989. See also McCulloch, 1995, and Rowlands, 1999.) Traditionally, the philosophy of mind has been almost exclusively "internalist"—that is, the workings of the mind, our thought processes, have been seen as acting entirely within the physical boundaries of the brain and skull. Internalism has been an implicit but essential component of the mind–body dualism most strongly associated with Descartes, and is still, in various forms, fundamental to many current models of cognition. Externalism, on the other hand, does not place the boundaries of cognition within the skull, but argues that there are many external facilities or processes that are necessary for cognition. Wittgenstein, who can be said to have had Externalist leanings, gave the example of our using a pencil and paper when we perform calculations. Many of us need such external implements for even simple calculations, but all of us need them for complex calculations. If we do not have a paper and pencil handy, we, quite literally, *cannot think*—the pencil and paper become a *sine qua non* for thought itself. Today, we have many such tools essential for thought: computers, databases, graphical plotters, and so on. None of us remembers everything he or she needs to conduct our daily affairs. Books, databases, and personal computers become necessary extensions of our memories. Without these implements, we would not be able to think the way we do.

The beginnings of Externalism, as a distinct movement in the philosophy of mind, finds its roots in Putnam's (1975) "Twin Earth" thought experiment. Putnam asked us to imagine that there was a "Twin Earth" that was exactly like our own earth, even to the point of having a "twin" of every person on this earth. But there was one aspect of Twin Earth that was different: On Twin Earth they had a substance they called "water," which was exactly like our own water except that instead of having a chemical structure H_2O it had a different structure that Putnam called "XYZ." Except for the different chemical structure, Twin Earth water had exactly the same function there as it does here: Twin Earthers

drank it, washed in it, poured it on their plants, and used it in squirt guns for amusement. Twin Earth "water" came out of the sky in the form of rain, and large amounts of it formed rivers, lakes, and oceans, just like ours does. Since the Twin Earthers' use of their "water" was exactly like our own use of water, their conception of water—that is, their idea of what it was and how it was used—was exactly the same as our own idea of what we called "water." In other words, what average Twin Earthers had "in their heads" about water was exactly the same as what we had in our heads about our version of water. Yet, Putnam wrote, Twin Earth water was different from our water because it had a different chemical structure (XYZ vs. H_2O). The ineluctable conclusion of this thought experiment is that semantic meaning is not entirely internal. At least part of the definition of what water is, is external to our skulls because what we and the Twin Earthers have in our heads cannot distinguish our water from Twin Earth water. As Putnam (1975, pp. 144) put it, "Cut the pie any way you like, 'meanings' just ain't in the head!" Tyler Burge (1979) published an article a few years later extending Putnam's externalist interpretation of semantics to include intentional mental states such as beliefs, desires, hopes, and fears. Burge called the internalist interpretation that he and Putnam criticized "individualism."

Although the Twin Earth thought experiment is entirely fanciful, similar phenomena occur every day. In most categories there is a level of generality where different people will call different things by the same name—for example, what I call a "sparrow" and another person calls a "sparrow" might actually be different species of birds, even though they have the same behavior, general appearance, and habitat. The Twin Earth thought experiment has had a profound effect on philosophy over the last three decades. As Pessin and Goldberg (1996, p. xi) observed, "Twin Earth and 'The Meaning of Meaning,' the article in which it became famous, comprise perhaps the most influential single philosophical episode in the past half century." This quotation was taken from the preface of their valuable 20-year retrospective collection of prominent articles written about Putnam's Twin Earth thought experiment.

Externalist theories of cognition have been appearing in areas outside of philosophy, too. Andy Clark (1997), a neuroscientist/philosopher, has extended it to cognitive science with what he calls "scaffolding." Scaffolding is the process in which people intentionally alter their

environment in ways that will elicit the kinds of responses that they want. Scaffolding provides external augmentation for intelligent activity, enabling us to achieve outcomes that would be difficult or impossible for a single, unassisted individual. This external assistance can be physical (e.g., a hammer, a truck, a boat, databases), cognitive (e.g., books, methods of estimation, rules of thumb, explicit directions), or social (e.g., creating professional societies or guilds of craftsmen to establish professional standards, to facilitate the dissemination of information, and to monitor professional conduct). Scaffolding contrasts most specifically with the Internalist foundation for intelligent behavior—"mental models." For some Externalists, mental models are not the foundation of understanding at all. As they put it, we don't need internal representations of the world, that is, mental models, because "the world is its own best representation" (Clark, 1997, p. 46).

Clark traces the roots of the idea of scaffolding back to the Soviet psychologist Lev Vygotsky (1986). As Clark (1997, p. 45) describes it: "Vygotsky stressed the way in which experience with external structures (including linguistic ones, such a words and sentences ...) might alter and inform an individual's intrinsic modes of processing and understanding." The more general notion of "mind as inextricably interwoven with body, world, and action" (Clark, 1997, p. xvii) has its antecedents in the works of philosophers Martin Heidegger (1961) and Maurice Merleau-Ponty (1962, 1963). This ability to alter our immediate environment in order to augment our abilities and stimulate specific actions gives us the capability to perform exceptionally complex tasks, from building a house, to constructing a dam, to designing the equipment that can take astronauts to the moon and return them safely to Earth. Scaffolding occurs even on a simple level when we make subtle changes in our environment, for example, to help us remember things; that is, we can put an overdue library book on the driver's seat of the car so that when we get into the car next we will see the book and be reminded to return it. Or, we can leave notes to ourselves stuck to prominent places, like the refrigerator, to remind us of things we need to do.

Some of the most interesting scaffolding is that which we erect to enable several individuals to work together to perform a complex task. An exceptionally rich and detailed example of this kind of deliberate scaffolding occurs in Hutchins' (1995) *Cognition in the Wild,* in which the

author describes a long and detailed study of the process of navigation on a Navy ship. This example is interesting because it involves the collaboration of several individuals, each of whom brings a different kind of expertise to the activity, and it requires a kind of precision and low fault tolerance that puts significant pressure on the individuals involved to get all the procedures right. As Clark (1997, p. 214), talking about Hutchins' work, put it, "Ship navigation emerges from the well-orchestrated adaptation of an extended complex system comprising individuals, instruments, and practices."

Scaffolding and Information Retrieval

If scaffolding can be considered an often necessary part of our cognitive processes, then it takes no great exercise of insight to see that information retrieval systems can be part of the scaffolding for many of our intelligent activities. An information retrieval system is, most obviously, a kind of external memory that can greatly augment what we remember, allowing us to consider and compare much more information than we could keep in our heads. But, more subtly, it can influence *how* we think as well. The particular searching procedures and the explicit or implicit theory of representation used by an information retrieval system can, quite literally, become extensions of the cognitive processes of inquirers—this can be either good or bad. For example, a simple full-text document retrieval system works by having searchers specify the words and phrases that they believe will occur in the literal text of the documents that have the intellectual content they would find useful, but would not occur in the text of documents they would not find useful. It has been shown, though, that on a reasonably large system, searchers looking for documents with a particular intellectual content are not very good at predicting the words and phrases that would occur in the documents they want, but would not occur in similar documents that they would not want (Blair & Maron, 1985). A simple full-text retrieval system, as scaffolding, extends the cognitive processes of the searcher, but it does so in an unnatural way, forcing the searcher to predict the exact words and phrases that occur in the desired documents, but do not occur in undesirable documents—something that people don't do well on a reasonably large system. Because information retrieval systems are potentially part

of the scaffolding of inquirers, and, thus, are extensions of their cognitive processes, it becomes important that the systems provide a good fit between what they do and what people are good at. In the example just discussed (Blair & Maron, 1985), the searchers were lawyers trying to find evidence that supported the defense of a large corporate lawsuit. Full-text retrieval was not a good extension of the lawyers' thought processes in this case because it is hard to predict the exact words and phrases that can be used to discuss a particular topic. But one could imagine situations where full-text retrieval would be a good extension of our cognitive processes: perhaps a system providing access to newspaper articles where most of the searches are for articles discussing specific individuals, companies, institutions, cities or countries, or within certain time frames. People *are* quite good at remembering proper names and approximate time frames. An article discussing Henry Kissinger's 1972 talks with Andrei Gromyko on East Berlin will, almost certainly, have each of these names and the date in the article, and articles that do not discuss this event will almost certainly *not* have all of these names in the literal text. Here, a full-text information retrieval system will extend the cognitive capacities of the searchers in ways that take advantage of how they think; that is, it facilitates, or even improves, their thinking. Naturally, information retrieval systems should augment what we *don't* do well, too, such as having literal recall of gigabytes of written text or images.

If IR systems can be like extensions of our thought processes, then we must take heed of the way that memory works. Human memory is not a faculty in which everything perceived is deposited literally and kept. Psychologists have shown that we are quite selective about what we remember, and time has a natural way of weeding out memories that are less important or less useful to us. In other words, forgetting may be just as important as remembering. "The Russian neuropsychologist Alexander Luria described a much-celebrated mnemonist, Shereshevskii, who ... was overwhelmed by detailed but useless recollections of trivial information and events. He could recount without error long lists of names, numbers, and just about anything else that Luria presented to him. ... Yet when he read a story or listened to other people, he recalled endless details without understanding much of what he read or heard ... he had great difficulty grasping abstract concepts" (Schacter,

1996 referring to Luria, 1968, p. 81). The literal, nonforgetting memory of Luria's mnemonist is probably a close analogy to the way that computer "memory" works, and the difficulties Shereshevskii had with total recall are certainly a caveat of sorts. Thus, if the ranking of memories by their importance and the forgetting of useless ones are significant processes in human recollection, then IR systems, if they are to be adequate scaffolding for intelligent activities, may need to have similar characteristics—records should be continually ranked by their importance and less important ones regularly weeded out and forgotten. IR systems do not always have to mimic the way human memory works, but they should complement its functionality. Sometimes this may mean doing things essentially the same way, but sometimes it may mean doing things very differently. This is one of the main points made by Winograd and Flores (1987).

The only article in the information science literature to make explicit use of Clark's work on scaffolding is Jacob's (2001). In this article, Jacob relates classification theory to the "everyday world of work," a context that is strongly reliant on scaffolding. Scaffolding theory will undoubtedly have an important impact on description and classification in information retrieval, so it is likely that there will be more publications utilizing it in the future.

Implications of the Philosophy of Language for Information Retrieval

The general features of the philosophy of language's theory of meaning that I have presented here can be summarized as follows:

1. "Meanings" are not linked to words (Wittgenstein).
2. "Meanings" are not concepts or any other single *thing* (Wittgenstein, Putnam).
3. To understand the meaning of a word is not to have some definition in your head, but to be able to use the word correctly in the activities and practices in which it is normally used (Wittgenstein). To understand a word means to know *when* to use it—which activities and practices

(Wittgenstein's "forms of life") it is relevant in—and *how* to use it (Wittgenstein's "language games").

4. "Let the use of words teach you their meaning" (Wittgenstein, 1953, p. 220). Meaning is not the same as use, but *emerges* through use.

5. Context is important for language. We often understand the situation in which language is used before we understand the words used (Wittgenstein). Meaning, in part, is an *external* notion—what we have in our heads, our ideas, are neither necessary nor sufficient for determining what we mean: Context and circumstances are often essential determinants of meaning (Putnam, Wittgenstein).

6. We make a variety of assumptions about the intentions of those with whom we talk. In particular, unless given evidence to the contrary, we assume that the individuals with whom we talk will cooperate with us and follow Grice's maxims.

Although I have presented the principal themes of several prominent philosophies of language here, the reader should keep a number of caveats in mind. First of all, these are not the only issues engaged by philosophers of language. The literature of the philosophy of language is both deep and extensive, ranging far beyond the intellectual boundaries of the issues presented here. Secondly, the philosophical conclusions presented here are by no means accepted by all philosophers. Philosophy of language, like most other active intellectual processes, remains very much a dialectic. Nevertheless, the philosophies presented here *are* prominent, and have established themselves as major landmarks in the intellectual landscape of the philosophy of language.

The Significance of the Philosophy of Language for Information Retrieval

Because the thesis of this discussion is that the philosophy of language has some significance for the problem of description in information retrieval, I will briefly sketch some of this significance.

1. If the contexts of activities and practices are important for under-
   standing language, it stands to reason that activities and practices
   are important for understanding document descriptions, too. As a
   consequence, it is essential for information retrieval systems to be
   as close as possible to the activities and practices that they serve.
   As Wittgenstein (1953, p. 220) said, we need to "let the use of words
   teach [us] the meaning." If we want to know what the descriptions
   used to represent a document *mean,* we must examine how these
   descriptions are used in the activities and practices that use that
   information—how do people ask for this information and how do
   they talk about it? One of the consequences of computerizing infor-
   mation retrieval systems is that the information they contain is
   often separated from these relevant activities and practices. Paper-
   based information has some obvious disadvantages regarding stor-
   age and copying when compared to the same information in
   electronic form. But paper-based information has one distinct
   advantage over electronic information: Because a paper document
   does not need delicate electronic equipment to present it, it can be
   carried and used almost anywhere—from the office, to the home, to
   a bus, to a rainy construction site, and so on. It is also easy to mark
   up, annotate, or highlight, and parts of it can be clipped out or pho-
   tocopied and distributed. Further, small accidents such as dropping
   the information or spilling coffee on it do not render paper unread-
   able, although information on a laptop could not stand such abuse.
   Consequently, paper-based information can remain close to the
   activities that produce or use it, and these activities can provide an
   interpretive context for that information. But when that informa-
   tion is computerized, the very act of computerization may have the
   effect of removing the information from the activity context that
   provides much of its meaning and interpretation. The importance
   of the proximity of information systems to the activities and prac-
   tices they serve was a major concern of Blair (1990).
2. If information retrieval systems cannot be physically near the
   activities and practices they support, then it may be useful to bring
   some of this context into the descriptions of the documents them-
   selves. This enhancement could be done by linking documents to
   the respective activities and practices in which they might be used,

and weeding out information germane to activities that have concluded. The first step in information retrieval system design, then, is to develop a detailed taxonomy of the various activities and practices that produce or use the information on the system. Each document on the system must then be explicitly linked to one or more of these activities or practices. For private industry, documents could be linked specifically to the value-creating activities and the core competencies of the firm (Prahalad & Hamel, 1990).

3. If information retrieval can be seen as a kind of conversation between the searchers and those who designed the system or represented the documents, then the quality of retrieval is in some sense related to the quality of this conversation. But one of the characteristics of conversation is that the conversing parties are able to respond immediately to each other's expressions. This immediate conversational feedback minimizes the number of misunderstandings that can occur, by allowing the conversationalists to clarify any confusions or ambiguities of meaning that might arise. But one of the principal characteristics of information retrieval systems, especially computer-based ones, is that they inevitably create a distance between the conversationalists—searchers and system designers/indexers—that prevents them from getting the immediate feedback so characteristic of normal conversation (Blair, 1990). Linguistic meaning emerges through the interaction of individuals trying to make themselves understood as they conduct their daily affairs. But because no immediate feedback or chance at clarification takes place for searchers using an information retrieval system, the local interactions from which meanings emerge just don't occur. Looking at the information retrieval process as a conversation helps to clarify the role of the professional searcher in the search process. The professional searcher (for example, a librarian) must assume, primarily, the role of an interpreter. He or she must interpret or explain the intended meaning of the document descriptions to the searcher, and help the searcher express information needs in ways that, when used as queries, will retrieve any useful documents that are available on the system.

The unnatural conversations between searchers and systems designers/indexers point to important avenues of research and system design. First and foremost, system designers should look into designing information retrieval systems that afford the opportunity for searchers and systems designers/indexers to converse in a more real-time mode. For systems in which this is impossible, it would be useful to develop procedures that use searcher feedback to adapt document descriptions. User feedback has been an area of IR research for decades, but no major commercial systems use these techniques, and, although considerable research into adaptive systems has been conducted, no real consensus has emerged about which techniques are the best or even which techniques are better than others. (For early work in adaptive information retrieval see Salton [1989]. Some of the most interesting work on adaptation uses genetic algorithms. An excellent discussion of its importance in IR can be found in Serich [1999].)

4. Because much of our intelligent activity is heavily scaffolded, it stands to reason that information retrieval systems may often be an integral part of that scaffolding. For scaffolding to work well, though, it must supplement, support, or extend our actual cognitive processes. In other words, information retrieval systems must be designed, at least in part, to work with some specific cognitive ability or process that is endemic to humans and is essential to our ability to search for information. The information retrieval problem will probably not be addressed satisfactorily if it is seen as a purely technical problem—that is, if the retrieval problem is addressed by simply taking advantage of specific technical resources or efficiencies, such as storage capacity or physical access rates. The danger here is that by designing systems that take advantage of certain technical resources or efficiencies, we may actually force searchers to act in unnatural or problematic ways (see our example of using simple full-text retrieval to make fine discriminations of intellectual content in large systems [Blair & Maron, 1985]).

5. Finally, it is becoming increasingly important to understand how the growth in the size of information retrieval systems affects the prospects for designing effective systems. This situation is not a direct consequence of the philosophy of language, but arose in my

initial discussion of the ways in which document descriptions can fail. Because it is cheaper to keep all electronic documents than to regularly weed out those that are no longer useful, most computer-based information retrieval systems will continue to become larger. This increase in size will alter the way that documents and search requests are represented, changing the primary strategy of document representation from description to discrimination. But theories of document representation are primarily oriented toward description and rarely take into consideration how well index terms discriminate. Most existing automatic indexing procedures used by commercial systems operate solely within the textual boundaries of the document they are indexing and make no allowances for how discriminating the assigned index terms are for actual searches on a particular system. The notion of "term discrimination" considered here is not just a comparison of term frequency occurrences, in which a term that occurs in just one document in the collection is considered a good discriminator and a term that appears in all the documents is not. What I mean by a good discriminator is a term that discriminates useful from useless documents for a typical searcher. So term discrimination, the way that I mean it, must take into consideration the searching characteristics, techniques, and judgments of the typical searcher using the system in addition to the frequency of occurrence of the term. For example, if you are looking only at term frequency as a basis for discrimination, then identifying each document in a collection by a unique accession number is an excellent discriminator. But from the searcher's point of view, the unique accession number by itself would not be a very useful discriminator because there may be no simple way to relate it to what the searcher would find useful.

As information retrieval systems grow larger, the pressure to discriminate useful from useless documents will become greater and the ability to discriminate will most likely get worse; mutatis mutandis, it is easier to discriminate, that is, to find, two useful documents among 10 useless ones, than it is to discriminate or find two useful documents among 1,000 useless ones. Thus, as document retrieval systems become larger, retrieval effectiveness will generally get worse (or retrieval effort for the same level of effectiveness will get greater, which amounts to the

same thing). Part of this problem can be mitigated by setting up rigorous document retention policies and other procedures for weeding out useless documents from existing systems. But some systems, such as the World Wide Web, will grow without any upper bound, so the exclusion of useless Web sites, even if it could be done systematically, would probably not be enough to mitigate the problems brought on by the growth in the overall number of Web sites. But for those systems for which periodic document weeding would be useful, the criteria for removing documents may be purely pragmatic: Documents should be weeded out of a system when the activities to which they are relevant have concluded.

There is a semantic lesson here, although it is a subtle one. One of the most important claims that Wittgenstein made in his philosophy of language was that questions of meaning in language cannot be adjudicated by appealing to abstract principles of semantics or to dictionary definitions, but can be resolved only by appealing to the ordinary usage of language. In short, whatever meaning there is in language, it is a meaning that emerges only from the day-to-day interactive usage of language—there can be no "better" or "more accurate" meaning in language than the meaning of ordinary discourse (this is why Wittgenstein's philosophy of language is called "ordinary language philosophy"). If the final criterion for semantics is everyday usage, then we can see very quickly that information retrieval systems are forcing us to use language in an unnatural way. Specifically, our language was never intended to be used to discriminate the intellectual content of small numbers of documents from vast numbers of other documents with similar intellectual content. In our typical day-to-day interactions we simply don't make such fine distinctions.

In the majority of information retrieval situations, as has been shown, the strategy for representing intellectual content is oriented almost exclusively toward the description of content rather than the discrimination of content. In one kind of information retrieval, however, discrimination is taken into account in the representation of intellectual content—cases in which the retrieval of information is a prominent part of an established practice. Good examples include some of the scientific disciplines, particularly the natural sciences. Here, the development of a taxonomy, which both describes and discriminates the major topics within the practice, is an important part of the practice itself. Biologists

expend a significant amount of time discussing and even arguing about how the plants, animals, and professional activities of their respective fields are to be represented—that is, how they are to be described and discriminated from other, often similar, elements in their field. In fact, the theory behind the development of taxonomies in the life sciences has become a field of study in itself. As document collections grow to unprecedented sizes, perhaps we will need to take a lesson from the taxonomical efforts of the natural sciences and find ways to develop taxonomies of the more mundane information that we deal with on a daily basis. To expect that there will be a simple technical solution to this problem—the development of a particularly fast search engine, for example—is to ignore the complexity of language to which the field of the philosophy of language is a testament.

Writings on the Philosophy of Language and IR

Although the relevance of the philosophy of language to information retrieval is significant, its actual direct impact on the IR literature has been modest. Some of this may be due to the difficulty of the philosophy literature, but it is also the case that the philosophy of language is primarily concerned with puzzles of its own—puzzles such as the boundaries of factual discourse or the supervenience of psychological states on brain states—which are of less obvious value for understanding the problems of document representation and retrieval. Consequently, the reader interested in information retrieval problems might have to read a fairly large body of writings before he or she could distill something useful from it. It is to be hoped that this discussion will have provided the reader with some useful entrées to that literature. Nonetheless, a few authors interested in the problems of information retrieval have found the philosophy of language useful. Frohmann (1990) utilized the philosophy of language's critique of mentalism to point out similar shortcomings in the mentalism of indexing theory. In the same year, Blair (1990) published *Language and Representation in Information Retrieval,* in which he presented an extended argument for the importance of the problem of representation in information retrieval and for the relevance of Wittgenstein's philosophy of language for understanding it better. Two

years later, Blair (1992) published a paper in which he gave a brief overview of the relevance of the linguistic philosophies of Austin, Searle, Grice, and Wittgenstein to the problems of information retrieval. In 1998, Hjørland published a paper in which he used the early and late philosophies of Wittgenstein as frameworks through which to examine some of the issues of information retrieval. In particular, he contrasts the early and late theories of meaning articulated by Wittgenstein in his "picture theory of meaning" and his "language games." Blair and Kimbrough (2002) have applied Wittgenstein's notion of "perspicuous examples" ("Übersichtliche Darstellungen") to the description of documents. They propose that in many document collections there are what they call "exemplary documents," which provide a guide to the intellectual content of many of the documents. Finally, there is an indirect link between an article on "relevance" by Harter (1992) and Grice's philosophy of language. Harter discusses the notion of relevance as presented by Sperber and Wilson (1986) who, in turn, base much of their work on Grice's philosophy of language.

One branch of the philosophy of language that has had a clear influence on information system design in general and, to a lesser extent, information retrieval has been the theory of Illocutionary, or Speech, Acts: Austin (1962) called them "Illocutionary Acts," while Searle (1969) gave them their more common name, "Speech Acts." Their biggest impact has been on electronic messaging systems (Kimbrough, 1990; Kimbrough & Lee 1986; Winograd & Flores, 1987), but Blair (1990, 1992) has described how they could be applied to information retrieval. The essence of Illocutionary, or Speech, Acts is that a class of linguistic events (Speech Acts) exists that has predictable structures and processes. I can say:

1. I'll mail you the check tomorrow.
2. I christen this ship the "Norton Sound, AVM-1."
3. Pick me up outside the main gate after the game.
4. Mary is the best copyeditor we have.

When we say such things (make a promise, christen a ship, give an order, or make an evaluation), we aren't so much talking *about* something, or making an assertion, we are actually doing something with our

statements. If a reasonable individual in a normal situation promises to do something, then, by virtue of that statement, he or she has made a promise. What makes a Speech Act work is a set of "felicity conditions" that must be satisfied. Felicity conditions are ordinary circumstances or conventions that each Speech Act presupposes. I can promise you that I will lend you my car, if I have a car. But I cannot promise to make you a member of the House of Lords, because I am not the ruling British Monarch. Searle (1969) identifies the following kinds of Speech Acts:

- Directives: In which we order others to do things (e.g., "Get me a Guinness Stout and a bag of chips.")
- Commissives: In which we promise to do something (e.g., "I'll return the book I borrowed tomorrow.")
- Declarations: In which we bring about changes in the world solely by our utterance—in short, "Saying makes it so" (e.g., "I now pronounce you husband and wife.")
- Expressives: In which we express our personal feelings and attitudes (e.g., "You did a terrific job!")
- Assertives: In which we make statements, truly or falsely, about how things are (e.g., "The Sears Tower is the tallest building in Chicago.")

Each of these acts has a predictable structure and felicity conditions that guarantee its success. The best known applications of Speech Acts to electronic messaging is the COORDINATOR system (Flores, Graves, Hartfield, & Winograd, 1988; Winograd, 1988; Winograd & Flores, 1987). In this system, if you send a message making a promise to someone, the COORDINATOR system will prompt you in the future to fulfill your promise (the COORDINATOR cannot tell whether you have made a promise, given an order, or made a declaration; nor can it ascertain whether all the felicity conditions have been satisfied—this information must be provided by the individual sending the message). The idea that a number of messages can be related as parts of the same transaction is an important consequence of Speech Act analysis.

Blair (1990, 1992) has suggested using the different Speech Acts as a way of categorizing messages on an information retrieval system. This would provide another kind of access for retrieving documents. Such a

classification system is especially useful for business communication in which the type of document—promise, order, declaration, and so on—can be important in many business processes (e.g., a "promise" in a business context might be a contract, whereas a statement of corporate strategy might be a kind of directive).

Finally, a number of authors in the information retrieval literature have found some use for an area of philosophy that historically has often been part of the philosophy of language—formal logic. Aristotle's syllogistic logic is arguably the first serious philosophy of language, and over the subsequent centuries logic has been primarily used to model formal relationships in language, such as the structure of argument and the nature of propositions. The early work in applying logic to the problems of IR was by Cooper (1971). More recently there has been a contribution by van Rijsbergen (1986a, 1986b, 1989), followed by a compilation by Crestani, Lalmas, and van Rijsbergen (1998). Crestani et al. bring together a nice selection of papers on logic and information retrieval. Cooper's paper defined a logically rigorous notion of relevance in information retrieval. The idea of relevance has been the focus of much concern and extensive writing, but little agreement. For Cooper, a document is relevant to a search query if its text can be used to form a "minimal premise set" that could be used to prove the assertion implicit in the search query. Although such a formal notion of relevance has a fairly narrow application, it does a useful job of establishing the boundaries of logical analysis in information retrieval. Some works, such as Wilson's (1973) writing on what he called "situational relevance" and van Rijsbergen's (1979), acknowledge an explicit debt to Cooper's seminal paper.

Conclusion

This discussion has attempted to provide an overview of some of the main ideas in the philosophy of language that have relevance to the issues of information retrieval. The philosophy of language is a much deeper and broader field of study than could be demonstrated in this chapter. Nevertheless, many of the most obvious connections between the philosophy of language and information retrieval should be apparent, and the relevance of the philosophy of language to the problems of

information retrieval should be evident. Much good work applying the insights of the philosophy of language to information retrieval remains to be done.

Bibliography

Ambrose, A., & Macdonald, M. (Eds.). (1979). *Wittgenstein's lectures: Cambridge, 1932–1935*. Chicago: University of Chicago Press.

Austin, J. L. (1962). *How to do things with words*. Oxford, UK: Oxford University Press.

Baker, G. P., & Hacker, P. M. S. (1985). *Analytical commentary on the philosophical investigations: Wittgenstein: Rules, grammar and necessity*, Oxford, UK: Basil Blackwell.

Baker, G. P., & Hacker, P. M. S. (1980). *An analytical commentary on Wittgenstein's Philosophical Investigations: Vol. 1. Wittgenstein, understanding and meaning*. Chicago: University of Chicago Press.

Black, M. (1968). *The labyrinth of language*. New York: New American Library.

Blackburn, S. (1984). *Spreading the word: Groundings in the philosophy of language*. Oxford, UK: Clarendon Press.

Blair, D. C. (1992). Information retrieval and the philosophy of language. *Computer Journal, 35*, 200–207.

Blair, D. C. (1990). *Language and representation in information retrieval*. Amsterdam: Elsevier Science.

Blair, D. C. (1986). Indeterminacy in the subject access to documents. *Information Processing & Management, 22*, 229–241.

Blair, D. C. (1980). Searching biases in large, interactive document retrieval systems. *Journal of the American Society for Information Science, 31*, 271–277.

Blair, D. C., & Kimbrough, S. O. (2002). Exemplary documents: A foundation for information retrieval design. *Information Processing & Management, 38*, 363–379.

Blair, D. C., & Maron, M. E. (1985). An evaluation of retrieval effectiveness for a full-text document retrieval system. *Communications of the ACM, 28*(3), 289–299.

Bouwsma, O. K., Wittgenstein, L., Craft, J. L., & Hustwit, R. E. (1986). *Wittgenstein: Conversations, 1949–1951*. Indianapolis, IN: Hackett Publishing Company.

Brooks, T. A. (1993). All the right descriptors: A test of the strategy of unlimited aliasing. *Journal of the American Society for Information Science, 44*, 137–147.

Burge, T. (1979). Individualism and the mental. In P. A. French, T. E. Uehling, Jr., & H. K. Wettstein (Eds.), *Midwest Studies in Philosophy, 4, Studies in Metaphysics* (pp. 73–122). Minneapolis: University of Minnesota Press.

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

Clark, A. (1997). *Being there: Putting brain, body, and world together again*. Cambridge, MA: MIT Press.

Cooper, W. S. (1971). A definition of relevance for information retrieval. *Information Storage and Retrieval, 7*, 19–37.

Crestani, F., Lalmas, M., & van Rijsbergen, C. J. (1998). *Information retrieval: Uncertainty and logics: Advanced models for the representation and retrieval of information*. Boston: Kluwer Academic.

Devitt, M., & Sterelny, K. (1999). *Language and reality: An introduction to the philosophy of language*, 2nd ed. Cambridge, MA: MIT Press.

Eco, U. (1976). *A theory of semiotics*. Bloomington, IN: Indiana University Press.

Fadiman, C., & Bernard, A. (Eds.). (2000). *Bartlett's book of anecdotes*. New York: Little Brown.

Flores, F., Graves, M., Hartfield, B., & Winograd, T. (1988). Computer systems and the design of organizational interaction. *ACM Transactions on Office Information Systems 6*, 153–172.

Frege, G. (1952). On sense and meaning. In P. Geach & M. Black (Eds.), *Translations from the philosophical writings of Gottlob Frege* (pp. 56–78). Totowa, NJ: Rowman and Littlefield.

Frohmann, B. (1990). Rules of indexing—A Critique of mentalism in information-retrieval research. *Journal of Documentation, 46*, 81–101.

Furnas, G. W., Landauer, T. K., Gomez, L. M., & Dumais, S. T. (1987). The vocabulary problem in human-system communication. *Communications of the ACM, 30*(11), 964–971.

Geach, P. T., Shah, K. J., & Jackson, A. C. (1989). *Wittgenstein's lectures on philosophical psychology, 1946–47*. Chicago: University of Chicago Press.

Glock, H.-J. (1996). *A Wittgenstein dictionary*. Cambridge, MA: Basil Blackwell.

Grice, P. (1989). *Studies in the way of words*. Cambridge, MA: Harvard University Press.

Hacker, P. M. S. (1989). *Insight and illusion: Themes in the philosophy of Wittgenstein*. Oxford, UK: Oxford University Press.

Hacker, P. M. S. (1990). *Analytical commentary on the philosophical investigations: Wittgenstein: Vol. 3. Meaning and mind*. Oxford, UK: Basil Blackwell.

Hacker, P. M. S. (1996a). *Analytical commentary on the philosophical investigations: Vol. 4. Wittgenstein: Mind and will*. Oxford, UK: Basil Blackwell.

Hacker, P. M. S. (1996b). *Wittgenstein's place in twentieth-century analytic philosophy*, Oxford, UK: Basil Blackwell.

Hallett, G. (1977). *A companion to Wittgenstein's Philosophical Investigations*, Ithaca, NY: Cornell University Press.

Harter, S. P. (1992). Psychological relevance and information science. *Journal of the American Society for Information Science, 43*, 602–615.

Heidegger, M. (1961). *Being and time*. (J. Macquarrie & E. Robinson, Trans.). New York: Harper and Row. (Original work published in 1927.)

Hjørland, B. (2000). Library and information science: Practice, theory, and philosophical basis. *Information Processing & Management, 36*, 503–531.

Hjørland, B. (1998). Information retrieval, text composition, and semantics. *Knowledge Organization, 25*, 16–31.

Holland, J. (1998). *Emergence: From chaos to order*. Reading, MA: Addison-Wesley.

Hutchins, E. (1995). *Cognition in the wild*. Cambridge, MA: MIT Press.

Jacob, E. (2001). The everyday world of work: Two approaches to the investigation of classification in context. *Journal of Documentation, 57*, 76–100.

Kenny, A. (1994). *The Wittgenstein reader*. Oxford, UK: Blackwell.

Kimbrough, S. O. (1990). On representing schemes for promising electronically. *Decision Support Systems, 6*, 99–122.

Kimbrough, S. O., & Lee, R. M. (1986). On illocutionary logic as a telecommunications language. *Proceedings of the 7th International Conference on Information Systems*, 15–26.

King, J., & Lee, D. (1978). *Wittgenstein's lectures: Cambridge, 1930–1932*. Chicago: University of Chicago Press.

Locke, J. (1985). An essay concerning human understanding. In A. J. Ayer & R. Winch (Eds.), *British empirical philosophers, (Book III)*. London: Routledge and Kegan Paul. (Original work published in 1690.)

Ludlow, P. (Ed.). (1997). *Readings in the philosophy of language*. Cambridge, MA: MIT Press.

Luria, A. R. (1998). *The mind of a mnemonist: A little book about a vast memory* (L. Solotaroff, Trans.). New York: Basic Books.

Lycan, W. (2000). *Philosophy of language: A contemporary introduction*. London: Routledge.

Malcolm, N. (1972). *Ludwig Wittgenstein: A memoir*. London: Oxford University Press.

Malcolm, N. (1986). *Nothing is hidden: Wittgenstein's criticism of his early thought*. Oxford, UK: Basil Blackwell.

Malcolm, N. (1995). *Wittgensteinian themes: Essays 1978–1989* (G. H. von Wright, Ed.). Ithaca, NY: Cornell University Press.

McCulloch, G. (1995). *The mind and its world*. New York: Routledge.

McGinn, C. (1989). *Mental content*, Oxford, UK: Basil Blackwell.

Merleau-Ponty, M. (1962). *The phenomenology of perception* (C. Smith, Trans.). London: Routledge and Kegan Paul.

Merleau-Ponty, M. (1963). *The structure of behavior* (A. L. Fisher, Trans.). Boston: Beacon Press.

Monk, R. (1990). *Ludwig Wittgenstein: The duty of genius*. New York: The Free Press.

Nedo, M. (Ed.). (1993). *Wiener Ausgabe* [Vienna edition]. Vienna: Springer-Verlag.

Pessin, A. & Goldberg, S. (Eds.). (1996). *The twin earth chronicles: Twenty years of reflections on Hilary Putnam's "The meaning of 'meaning.'"* Armonk, NY: M. E. Sharpe.

Pitkin, H. (1972). *Wittgenstein and justice*. Berkeley, CA: University of California Press.

Prahalad, C. K., & Hamel, G. (1990). The core competencies of the corporation. *Harvard Business Review, 68*(3), 79–91.

Putnam, H. (1975). The meaning of "meaning." In K. Gunderson (Ed.), *Minnesota studies in the philosophy of science*: Vol. 7. Language, mind and knowledge (pp. 131–193). Minneapolis, MN: University of Minnesota Press.

Putnam, H. (1988). *Representation and reality*, Cambridge, MA: MIT Press.

Redpath, T. (1990). *Ludwig Wittgenstein: A student's memoir*. London: Duckworth.

Rorty, R. (Ed.). (1967). *The linguistic turn: Recent essays in philosophical method*. Chicago: University of Chicago Press.

Rosenberg, J. F., & Travis, C. (Eds.). (1971). *Readings in the philosophy of language*. Englewood Cliffs, NJ: Prentice-Hall.

Rowlands, M. (1999). *The body in mind: Understanding cognitive processes*. New York: Cambridge University Press.

Russell, B. (1986). The philosophy of logical atomism and other essays, 1914–19. In J. G. Slater (Ed.), *The collected papers of Bertrand Russell: Vol. 8*. London: Allen & Unwin.

Russell, B. (1905). On denoting. *Mind, 14*, 479–493. [Reprinted in R. Marsh (Ed.). (1956) *Logic and knowledge*. London: Allen & Unwin.].

Ryle, G. (1931). Systematically misleading expressions. *Proceedings of the Aristotelian Society, 32*, 139–170.

Salton, G. (1989). *Automatic text processing: The transformation, analysis, and retrieval of information by computer*. Reading, MA: Addison-Wesley.

Schacter, D. (1996). *Searching for memory: The brain, the mind, and the past*. New York: Basic Books.

Searle, J. (1969). *Speech acts: An essay in the philosophy of language*. Cambridge, UK: Cambridge University Press.

Serich, S. T. (1999). Zipf's law as a necessary condition for mitigating the scaling problem in rule-based agents. (Doctoral dissertation, University of Michigan). (UMI No. 9959857).

Sperber, D. & Wilson, D. (1986). *Relevance: Communication and cognition*. Oxford, UK: Basil Blackwell.

Swanson, D. R. (1996). *Studies of indexing depth and retrieval effectiveness*. Unpublished report, National Science Foundation Grant GN 380.

Strawson, P. F. (1950). On referring. *Mind, 59*, 320–344.

Toyton, E. (1997, June). The Wittgenstein controversy. *Atlantic Monthly, 276*(6), 28–41.

van Rijsbergen, C. J. (1979). *Information retrieval*, 2nd ed., London: Butterworths.

van Rijsbergen, C. J. (1986a). A new theoretical framework for information retrieval. *Proceedings of the Annual International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR '86)*, 194–200.

van Rijsbergen, C. J. (1986b). A non-classical logic for information retrieval. *Computer Journal, 29*, 481–485.

van Rijsbergen, C. J., & Lalmas, M. (1996). An information calculus for information retrieval. *Journal for the American Society for Information Science, 47*, 385–398.

Vygotsky, L. (1986). *Thought and language* (A. Kozulin, Trans.). Cambridge, MA: MIT Press.

Wilson, P. (1973). Situational relevance. *Information Storage and Retrieval, 9,* 457–471.

Winograd, T. (1988). Where the action is. *Byte, 13,* 256A–258.

Winograd, T., & Flores, F. (1987). *Understanding computers and cognition: A new foundation for design.* Reading, MA: Addison Wesley.

Wittgenstein, L. (1953). *Philosophical investigations* (G. E. M. Anscombe, Trans.). New York: Macmillan.

Wittgenstein, L. (1958). *The blue and brown books.* New York: Harper.

Wittgenstein, L. (1961a). *Notebooks: 1914–1916* (G. H. von Wright & G. E. M. Anscombe, Eds.; G. E. M. Anscombe, Trans.). New York: Harper.

Wittgenstein, L. (1961b). *Tractatus logico-philosophicus* (D. F. Pears & B. F. McGuinness, Trans.). London: Routledge & Paul. (Original work published in 1921.)

Wittgenstein, L. (1967). *Zettel* (G.E.M. Anscombe & G.H. von Wright, Eds.; G. E. M. Anscombe, Trans.) Oxford, U. K.: Basil Blackwell.

Wittgenstein, L. (1969). *On certainty.* Oxford, U. K.: Basil Blackwell.

Wittgenstein, L. (1972). *Lectures and conversations on aesthetics, psychology, and religious belief* (C. Barrett, Ed.). Berkeley, CA: University of California Press.

Wittgenstein, L. (1974). *Philosophical grammar.* Oxford, U. K.: Basil Blackwell.

Wittgenstein, L. (1975). *Philosophical remarks.* Oxford, U. K.: Basil Blackwell.

Wittgenstein, L. (1978). *Remarks on the foundations of mathematics* (Rev. ed.) (G. H. von Wright, R. Rhees, & G. E. M. Anscombe, Eds; G. E. M. Anscombe, Trans.) Cambridge, MA: MIT Press.

Wittgenstein, L. (1980). *Remarks on the philosophy of psychology* (G.E.M. Anscombe, G.H. von Wright, & H. Nyman, Eds.; G. E. M. Anscombe, C.G. Luckhardt & M. A. E. Aue, Trans.) Chicago: University of Chicago Press.

Wittgenstein, L. (1992). *Last writings on the philosophy of psychology,* Vol. 2. *The inner and the outer, 1949–1951* (G. H. von Wright & H. Nyman, Eds.; C. G. Luckhardt & M. A. E. Aue, Trans.) Oxford, UK: Basil Blackwell.