Chapter 1

An Introduction to Methodology: Explanation, Specification, and Two Paradigms for Theorizing about Language

Language is a much explored but largely unknown territory. Expeditions into it are guided by overall visions of it. The Chomskian perspective takes the key to language to be its *infinite productivity*: the capacity to produce an infinite number of grammatical sentences from a finite basis. In current analytical philosophy, most scholars focus on the *compositional nature* of sentences, which allows the meanings of sentences to be a function of the meanings of their parts. Productivity and compositionality I find interesting, but a different feature of language amazes me: that noises can convey meaningful content. For instance, I use noises to enable a person to understand my claim that a certain tree is likely to fall within the next few weeks. Or I use noises to request someone to hand me a particular red book that I am pointing at. How can our noises convey such content? This question targets the relational nature of language: how do these noises of language produce relations to things in the world and to other people?

The question of how meaningful content is present in language is wholly different from that of productivity, in that the latter is an intrinsic property of a system, whereas meaningful content is a relational property between symbols in a system and things outside that system. Questions about meaningful content are broader than questions about compositionality. The latter is one feature of meaning, concerning part-whole relations in meaning; this relation does not directly bring in the connection of meanings to speakers and the world, as does the question of how words have meaningful content. I don't intend to rank these questions in importance, because they are heuristic questions, which cannot be ranked until the results of the explorations are in. Rather, I want to point out

that my question is different. It is also a traditional question, with traditional problems.

One problem with the question of how noises can convey content is that it raises issues about two relations at once, a relation to things in the world and a relation to people who understand the noises. Interestingly enough, the two traditional answers to the question are each based on a different one of these relations. One such answer is that meaningful content derives from the referential connection of words with things extrinsic to mind and language. Plato,[1] for instance, viewed meanings as based on the reference of words to eternal unchanging exemplars (the Platonic Forms) after which the material world is patterned. Later philosophers suggested that words might instead derive their meanings from referring to various other entities; among these were abstractions like universals or sets, or ordinary things in the material world. These theories get lumped together as *referential theories of meaning* because they all view meaningful content as arising from a referential relation between words and things. Gilbert Ryle (1957) caricatured such theories as the "'Fido'-Fido" theory: the meaning of the term 'Fido' is the dog Fido.

In contrast to referential theories, *ideational theories* view meaningful content in language as the result of a relation between words and ideas. On this view, a word gets its content from the ideas with which it is associated. Thus, ideas are taken to be intermediary devices linking words and things. *The Essay Concerning Human Understanding* by John Locke is probably the classical location for this theory, but versions of it circulated in medieval times and in the twentieth century, for example, in *The Meaning of Meaning* by C.K. Ogden and I.A. Richards. Also in the twentieth century, behaviorists developed a third theory of meaning. On this view, meaning is explained as a stimulus-response connection between the noises of language and behaviors: the noises are taken to be either stimuli for behavior, or responses to behavior.

All three theories of meaning encountered overwhelming objections (cf. Brown (1958) and Alston (1964)). I won't recount these, but will mention an interesting common feature in the trio. Underlying each theory is a fact about language, which was inflated into the essence of all language. For the *referential theory*, the fact is that words refer to things. For the *ideational theory*, it is that words are connected to ideas, and for the *behavioral theory*, it is that words are connected to behavior. These three facts cannot be ignored by any theory of how meaningful content connects with

noises, but none of them provides the basic essence of language. Instead they contain phenomena of language that must themselves be explained.

The question of how meaningful content is connected with language has been in disfavor for years, and has been replaced to a large extent by questions about language that are mostly specificatory (the nature of these will be discussed shortly). Although I still find the question about the nexus of meanings and language interesting, I think that the general form of this question that I have been discussing is too broad to answer. Meaning in language has many levels. Because of this, there cannot be a single uniform theory of meaning in general, at least at the beginnings of theories of meaning. I distinguish the following levels of meaning in language: (*a*) lexical meaning for individual words, (*b*) reference made by means of noun phrases and other referring expressions, (*c*) the meanings of sentences (these include two different kinds of meaning: propositional content and speech-act meaning such as asserting or questioning), (*d*) Gricean implicatures (meanings that are implied but not logically entailed), and (*e*) meanings in figures of speech.

The best way to begin a study of how meaningful content connects with language is, in my opinion, to pick one of these levels of meaning, and develop an explanatory theory for it. This is what I propose to do. I will study reference made by means of *referring expressions*.[2] Such reference lies between lexical meaning and sentence meaning, in that it presupposes and uses the former, and makes a contribution to the latter. For example:

1. That horse is fast.

To figure out what (1) means as used in a context, one must figure out the referent of 'That horse,' which in turn requires that one use the lexical meanings of 'that' and 'horse' plus other factors like gestures and the context to determine the reference of 'that horse.' For brevity, I will often speak of the reference of referring expressions, even though my models are for the entire process involved in using referring expressions; this process includes speaker, hearer, words, actions, and context.[3] Once the reference of 'That horse' is discerned, it is contributed in some way to the meaning of the sentence as a whole.

My exploration of reference concentrates on indexicals (terms like 'this,' 'I,' or 'now') and definite and indefinite descriptions. I leave aside proper names because I view their mechanisms as rela-

tively hidden, and best understood by comparison and contrast to other types of reference. Nor do I attempt to explain mechanisms underlying lexical meaning.

I picked the issue of the reference of referring expressions ('REs' for short) as a beginning to the study of the relation of words to things and thought, because REs link conversers with things in the world. In addition, reference is a central issue in twentieth-century philosophy, with well-defined problems and ample data. Some of these philosophical problems have been around since Gottlob Frege, the founder of the analytical tradition. For instance, how can we refer to the nonexistent, how are identity statements informative, or why is substitution of coreferential REs invalid in contexts of propositional attitude (contexts within the scope of verbs like 'believe that' or 'hope that')? There are still no generally accepted answers to these questions. Bearing on these problems about reference are data from communication, linguistics, and logical intuitions. Despite the importance of the problems and the availability of data, questions about reference to things in the outside world are relatively neglected in the Chomskian and compositional-semantic traditions. These schools do study reference, but concentrate on systems for representing reference and coreferentiality, rather than on how reference works in connecting words with things.

To answer the question of how reference works in connecting words and things, one needs an explanation. The view of explanation that I use is that of Rom Harré (1961), Jerrold L. Aronson (1984), and Fred I. Dretske (1981).[4] The goal, as Dretske (1981: 47) puts it, is "a more or less complete, precise, and systematic description of those entities and processes underlying the phenomena of interest." I devise models to clarify the underlying entities and processes which constitute reference. By a model, I do not mean a formal model, which is a set of objects that satisfy the axioms of a formal system. This type of model need not be explanatory, since it may use abstract data sets, rather than entities that resemble those underlying the phenomenon. Instead I have in mind a model that redescribes the phenomenon to be explained in terms of a set of entities which are simplified and abstract in comparison to real world entities, but still closely resemble the latter. For example, entities used in my models include the speaker, the hearer, the use of an RE, space-time locations, gestures, and descriptive content; also, relations of these entities are included in the models. These entities and relations closely resemble but are not identical to real

world entities described by the same terms. This non-identity is needed because the real world entities and relations are more complex than those in my models. But we need the simplicity of the models to begin to find explanations. I view my models for reference as first steps, undoubtedly in need of further refinement.[5] Nevertheless, the differences between the models and the real world are small enough that the models provide explanations that tie diverse phenomena together in a clear way, and lead to further explanations and predictions.

Explanatory models differ from what I will call 'specificatory theories.' Because this distinction is unclear in current theories of language, and crucial to my approach to reference, I will discuss it at length. Specificatory theories prevail, in my opinion, in current theorizing about language, though philosophers and linguists often speak of them as explanatory.[6] Specificatory theories aim at providing clear and systematic specifications of all the phenomena in an area of study. Such theories may be bases for precise predictions, but, because they need not use conditions underlying the phenomena, they need not provide explanations. In place of underlying conditions, specificatory theories may use stipulations, circular devices, ad hoc devices, parallel entities, or results of the phenomena. Ptolemaic astronomy, for example, was a successful specificatory theory, which was refined over centuries to provide moderately accurate specifications for movements of heavenly bodies. These specifications had important applications in agriculture and navigation. Specificatory and explanatory devices may be mixed together in a theory, and a fully mature science has explanatory theories which also provide precise specifications.

Examples from the history of science will help to clarify the distinction between specificatory and explanatory theories. For instance, Johannes Kepler's descriptions of the solar system were very accurate at specifying motions, but the underlying mechanism he offered (the *anima motrix* of the sun) was inadequate. Thus, Kepler provided good specifications but weak explanations. Isaac Newton's laws, which included universal gravitation, provided the needed explanatory mechanisms.

Another example is the classification of the elements by their spectographic lines.[7] Each chemical element was known to have its own distinct spectographic pattern for some time before the discovery of the conditions and mechanisms underlying this phenomenon. Thus, a specificatory theory associated spectographic data with each element, even though this association was unexplained.

Niels Bohr's model of the atom provided a key ingredient in the explanation; electron rings plus mechanisms for energy absorption and radiation explained the spectographic differences.

A specificatory theory need not provide an explanation even if it is descriptively complete, that is, for every interesting property and relation in the real world, there is a corresponding value of a variable in the specificatory theory, and there are laws describing how the values of the variables are related. Kepler's laws of motion and the classification of the elements by spectographic lines approximated descriptive completeness without being explanatory. This does not mean that these specificatory theories were unimportant. Indeed, they were outstanding scientific achievements. My point is that when we have specificatory theories, we still hunger for more: we want to know the conditions and mechanisms underlying the specifications.

The history of science shows interactions of specificatory and explanatory theories. In the examples given above, specificatory theories existed without accounts of underlying conditions. In other cases, a partial theory about underlying conditions exists with little specificatory detail (e.g., early heliocentric theories of Aristarchus, Philolaus, and Hicetas,[8] or the early theory of genes as carriers of heredity). Each type of theory can stimulate the development of the other.[9] But science requires both aspects in its theories, so that it is a mistake to rest content in either type by itself. Moreover, when explanatory mechanisms are found for specificatory theories, they often provide more precise specifications of the phenomena. For instance, Newtonian specifications of planetary orbits included the gravitational effects of planets on each other. And explanations of one set of phenomena often are useful for explaining additional phenomena, for example, Bohr's model of the atom and its electron rings was used not only in explaining the spectra of elements but also in explaining chemical bonding.

In recent work in philosophy of language, specificatory theories predominate.[10] These theories aim at laying out in a systematic way (by rules or axioms) the truth conditions for all sentences of a particular language (or of a fragment of a particular language). The extent to which these theories are also explanatory for sentence meaning is unclear, and I will not discuss it here. Behind these theories lies an approach to theorizing about natural language which I call the "*predicate logic paradigm.*"[11] Its current form has three features which affect the study of reference: it emphasizes (*a*) *translation into predicate logic*, and (*b*) *composi-*

tionality of meaning for sentences, and (*c*) it *separates semantics from pragmatics*. Each of these, I will argue, raises obstacles to developing explanatory theories of reference. Note that my focus is on the effects of predicate logic paradigm on *explanatory theories of reference*, and not on its merits generally.[12]

Translations of natural language sentences into predicate logic foster the compositional goal of the predicate logic paradigm because they provide a clear and systematic analysis of the combination of the meanings of the parts of a sentence into the meaning of the whole. Translation, however, aims at expressions equivalent to natural language sentences, and not at entities and mechanisms underlying those sentences. There is no reason to expect translations to achieve more than their explicit goal of equivalent expressions.[13] This point about translations does not depend upon the target language into which the translation is made: it doesn't matter whether the target language is matrix theory or Polish. That the target language is predicate logic, however, raises additional problems for explaining reference. Even though predicate logic functions mainly as a specificatory device within the predicate logic paradigm, it also supplies a mechanism used to explain reference, namely, predication.[14] Predication is used to define 'denotation': a term denotes whatever it is truly predicated of, or for singular reference, whatever it uniquely describes. Philosophers working within the predicate logic paradigm have used denotation to connect referents not only with REs that are definite or indefinite descriptions, but also with indexicals and proper names. But there are problems in using denotation as a mechanism of reference. Saul A. Kripke (1980) and Keith Donnellan (1970) have argued that proper names do not function as true descriptions of their referents.[15] In chapters 2 through 5, I argue that unique description is not the mechanism underlying either indexical reference or certain types of uses of definite and indefinite descriptions.

The focus on compositionality of meaning in the predicate logic paradigm also hinders the search for entities and mechanisms underlying reference. This focus makes truth conditions for sentences the paramount issue, so that questions about reference are treated in a top-down manner, that is, truth conditions are intuited for sentences, and the reference of an RE in them is simply whatever will help produce those truth conditions. This top-down approach ignores bottom-up considerations that are essential to an explanation: these include the actual functioning in communication of REs in relation to actions and context.

In addition to hindering the development of explanatory theories of reference, the top-down approach to reference makes specificatory theories of reference impossible. Arguments for the latter point have been given by two of the most prominent scholars working within the predicate logic paradigm, W. V. Quine and Hilary Putnam. Quine (1960, 1969) argues that reference is inscrutable on the grounds that the same truth conditions can be assigned to sentences despite differing assignments of referents to terms. This entails that reference is not specified by specifying truth conditions for sentences. Putnam (1981) develops a related argument for model-theoretic semantics. I take Quine's and Putnam's arguments to show that purely top-down theories which take truth conditions of sentences as the sole determiners of reference fail to specify reference. I propose that additional bottom-up theories may help to specify reference. And surely a bottom-up approach is needed for finding an explanatory model of the conditions underlying reference.

Yet another obstacle to the development of explanatory models for reference is the requirement that semantics is to be separated from pragmatics. These two are distinguished (in a tradition stemming from Charles W. Morris (1938) and Yehoshua Bar-Hillel (1954)) as follows: pragmatics treats the relations of symbols, objects, people and contexts, whereas semantics omits the latter two factors, and studies only the relations of symbols to objects. In practice, however, pragmatics is often defined negatively, as meaning which is not syntactical and not semantical. Like many negatively defined things, it has heterogenous components. It includes not only the reference of indexicals but also speech acts (e.g., the issue of whether a use of a sentence expresses an assertion, a command, or a question). In addition, it includes issues about communication generally, and about Gricean conversational implicatures (inferences based on rules of cooperation in conversation, which come into play after semantic content has been determined).

Associated with the distinction between semantics and pragmatics is an assumption that semantics is more important, and that pragmatics can be set aside while one works on semantics. Some philosophers apply this to reference, and assume that semantic reference can be treated independently of the pragmatics of reference, and that data from communication are not to be used for theories of reference.[16]

Are my models for reference a part of pragmatics? I don't view them in that way, because I do not accept the a priori distinction

between semantics and the pragmatics of reference. Before we study the phenomena of language empirically, we do not know either the content of the theories of language or the divisions of these theories. Ptolemaic astronomy would have been an adequate specificatory theory but a false explanatory theory if it was limited to the fixed stars. An a priori constraint which separates fields of study may get in the way of cutting nature at the joints.

The a priori separation of semantics and pragmatics affects not only the content of theories but also the availability of evidence. Communication is the main source of evidence for a bottom-up account of reference, but separating semantics from pragmatics restricts this evidence to pragmatics. This unavailability of bottom-up evidence for reference in semantics reinforces the top-down approach of truth-conditional semantics.

The discipline of artificial intelligence has moved away from the separation of syntax, semantics, and pragmatics because programs based on this separation did not run. One move was to adopt theories based on Charles J. Fillmore's (1968) case grammar. These theories use schemas or frames which combine syntactical and semantical features of meaning. Another move was to combine semantics with pragmatics, for example, by including within representations of meaning both default conditions based on what usually happens, and background or contextual knowledge.[17]

In philosophy, the separation of semantics and pragmatics has led to splitting apart linguistic and cognitive matters. This new dichotomy arises from current semantic theories of reference for indexicals and proper names, especially the direct reference theory, which takes certain types of REs to contribute only a referent to propositional content.[18] Because this restrictive view leaves reference with very little semantic content, some philosophers (e.g., Perry (1979, 1988) and Wettstein (1986, 1988) have concluded that cognitive features of reference are separate from semantic features. This new dichotomy is based on the old one between semantics and pragmatics, since the informative content of reference that goes beyond simply having a certain referent is ignored, shipped off to pragmatics, or both.

The three features of the predicate logic paradigm combine to promote certain a priori tendencies regarding theories of reference, especially in regard to reference as *social* and *perceptual*. The separation of semantics from pragmatics, by removing the study of communication from semantics, also removes the most obvious data about social and perceptual features of reference. The em-

phases on translations into predicate logic and on truth conditions also direct attention away from social and perceptual matters. And predicate logic suggests no likely representations for either social or perceptual features. In addition, a top-down approach that focuses on truth conditions for sentences is likely to miss the bottom-up role of perception in reference.[19] These a priori obstacles to seeing reference as perceptual and social are, in my opinion, of the highest importance, since I take the most basic cases of reference, and the simplest cases to model, to be both social and perceptual.

Even if I cannot use predicate logic to discover models for reference, couldn't I at least reexpress my theories in it? Saul A. Kripke (1980: 88, note 38), who refers to Robert Nozick in the following passage, seems to think so, at least for proper names:

> . . . There is a sense in which a description theory must be trivially true if any theory of the reference of names, spelled out in terms independent of the notion of reference, is available. For if such a theory gives conditions under which an object is to be the referent of a name, then it of course uniquely satisfies these conditions.

The point here is that one can use any theory of reference as a basis for conditions whose satisfaction indicates the referent of a name. This is incorrect for an explanatory theory of reference, because such a theory does not determine particular outcomes by itself, but only in combination with subsidiary theories which allow precise calculations of initial conditions.[20] In the sciences, explanatory models are usually found long before their precise applications to the relevant initial conditions are possible. For instance, we have great explanatory models for weather, but characterizing and computing precisely the relevant initial conditions are currently beyond us, with well known results for weather predictions.[21] Because of this, we cannot restate our theory of weather in terms of necessary and sufficient conditions for particular instances. Therefore, having a correct explanatory theory for reference does not imply that one can restate it in terms of necessary and sufficient satisfaction conditions for reference.[22]

If one could deal with initial conditions precisely, then one could rephrase a non-description theory of reference into a predication theory (one which works via true descriptions of the referent). However, such a theory would still have two extremely important limits. First, such re-expression of a theory will produce a theory that is specificatory and not explanatory. This is because (*ex hy-*

pothesi) the predicate logic devices used in the new theory do not play a basic role in the conditions underlying reference. Second, even if the predication theory can express a (specificatory) theory of reference that is already known, this does not show that the predication theory would be of any help in discovering yet unknown theories of reference. This heuristic point is of the utmost importance in my present work. In my opinion, predicate logic presents a heuristic obstacle to theories of reference; because it offers only one mechanism for reference, the focus of discussion shifts from how reference works in natural language to how reference is to be represented in the predicate logic notation. This orientation distracts one from imagining new models for reference.

If the predicate logic paradigm is not conducive to explanatory models for reference, what paradigm is? How can we get at the underlying conditions and mechanisms of reference? I use what I call the '*communicational paradigm.*' This paradigm has two central features: (*a*) the goal of devising *explanatory models*, and (*b*) an emphasis on *data from communication.* Explanatory models for reference require finding simplified abstracted versions of the entities and mechanisms which underlie (either constitute or cause) the phenomena of reference. Underlying conditions and mechanisms provide a bottom-up account of reference, and not merely a top-down account based on truth conditions of sentences. In laying out these underlying factors, the models abstract from and simplify the blooming buzzing confusion of natural language use. This simplification does not make the models less empirical, since all their ingredients are empirically supported. Such simplified models are likely first steps towards understanding how reference works.

To find underlying conditions requires more than finding truth-conditionally equivalent statements. Translation, then, because it aims at such equivalence, is not a useful tool in developing explanatory models. Because equivalence is not the goal, neither are analyses in terms of necessary and sufficient truth conditions. Nevertheless, explanatory models do yield analyses of the ingredients that enter into a phenomenon. Such analyses differ from truth-conditional analyses, and instead are like chemical analyses. Consider the parallel: water is chemically analyzed into hydrogen and oxygen, and table salt into sodium and chlorine. However, a list of the elements that go into water and table salt does not constitute their chemical explanation. Also required are the mechanisms by which the elements constitute the compounds: for water, the mechanism of chemical bonding is electron sharing between hydrogen and oxy-

gen, whereas for table salt, the mechanism is electrical attraction between sodium ions and chlorine ions. The mechanisms plus the chemical analysis constitute the chemical explanation.

The point of this parallel between chemical explanations and explanations of reference is twofold. First, the analyses that result from the explanatory models for reference provide ingredients that constitute the phenomenon, rather than expressions equivalent in meaning or extension to the phenomenon. Second, the analyses list the ingredients in reference, but the mechanisms by which those ingredients constitute reference are not part of the list, and instead are described elsewhere. Both the ingredients and the mechanisms are needed for an explanatory model.

Explanatory models must be based on appropriate data. Where are such data found for models of reference? The main source is acts of *communication*, in which speaker and hearer use words, gestures, and context to refer and to understand reference. Thus, the two central features of the communicational paradigm are closely connected: the goal of explanatory models requires a prominent place for data from communication.

An emphasis on communication has important corollaries which I view as naturally associated with the communicational paradigm, though not absolutely required by it. Foremost of these is an impetus toward a social-psychological model, rather than an individualistic one. The data from communication suggest ingredients for the models, and these data involve interactions between speaker and hearer. My social-psychological approach to reference fits well with Noam Chomsky's view that linguistics is a part of social psychology. The data from communication also suggest a role for physical objects in the models. The presence of conversers and physical objects in the models makes them fit well with Tyler Burge's (1979, 1986, 1990) recent arguments against individualism in philosophy of mind. Evidence from communication also suggests *perception* as a key ingredient in indexical reference.

Besides suggesting models, the communicational paradigm also fosters naturalism. Naturalism in philosophy has three requirements: an empirical rather than a priori approach to theorizing, the avoidance of other-worldly entities in hypotheses (e.g., Platonic forms, Fregean senses, or possible worlds construed as ontologically fundamental), and a coherence with the sciences in content and method.

Empirical data from communication have primacy of place in the communicational paradigm, and such data are large in quan-

tity, public, and not inscrutable. It seems obvious to me, as it does to Howard K. Wettstein (1989), that it is easier to start theorizing about reference with something public and social, rather than with private thoughts inside the mind.[23] Also, the communicational paradigm avoids a priori commitments to formalisms, or to separating semantics and pragmatics; in these matters it is relatively more empirical than the predicate logic paradigm. At the present stage of development for theories of reference, it would be happenstance to find a formalism that works for reference. In general, we cannot know which formal system is appropriate to a field of study until we have some knowledge of the important parameters in that field, and of relationships of those parameters.[24] The present study aims at finding some parameters and relationships for a theory of reference.

The communicational paradigm sustains naturalism in yet other ways. The entities hypothesized in my models closely resemble ingredients in communication, so that they are this-worldly rather than other-worldly. In addition, the communicational paradigm aims at continuity with the sciences in a general way by aiming at explanatory models, and in more particular ways by utilizing notions and findings from psychology, artificial intelligence, and linguistics.

Although the communicational paradigm is constrained to seek explanatory models, and to treat data from communication as primary, in most other matters it is wide open to data and to theories, including hypotheses based on predication, perception, or social interactions.

What would success look like under the communicational paradigm for reference? Its test for explanatory models is that of the sciences, namely, checking their explanatory and predictive power. Gilbert Harman's (1973) notion of the best *total* explanatory account is important here: these models should contribute to explaining phenomena found in communication and in logical intuitions for all linguistic contexts, as contrasted to explaining only a limited range of phenomena while omitting others as has often been done in philosophy of language. Among the phenomena set aside by various approaches to language are contextual dependencies of language, cognitive aspects of language, and certain linguistic structures, such as existence claims or contexts of propositional attitude. I am not claiming that I discuss all possible phenomena of reference, but rather that they are all relevant to the acceptability of my models.

If a model for reference can explain the phenomena of communication, logic, and linguistics which prompted its introduction, it has some success. If its explanatory powers extend beyond those phenomena, this is stronger support. In this regard, I will argue that the models I use to explain the referential functioning of indexicals and descriptions also provide explanations for traditional problems about reference such as reference to the nonexistent, the informativeness of identity statements, and referential opacity. Whether the models are also explanatorily fruitful for other kinds of phenomena, as in philosophy of mind, philosophy of science, artificial intelligence, psychology, or linguistics, only time will tell.[25] I take such fecundity to be the main mark of a good explanation.