Chapter 1

The Foundations of Performativity: Austin and Frege

Introduction

Austin's work on performativity is a crucial meeting point in the contemporary debates between philosophy and literary studies. As the translator of Frege's *Grundlagen,* Austin is heavily influenced by Frege's legacy, which enters British philosophy through Russell and Wittgenstein. The Derrida-Searle debates over the interpretation and implications of Austin's work demonstrate the English philosopher's continuing relevance. His analysis of the constative and locutionary dimensions of speech acts acknowledges the descriptive aspects of language, but his discovery of performative and illocutionary speech acts point to a new "condition of possibility" of language in which description itself becomes just one function of language. Although his analysis of speech acts uses Frege's sense-reference distinction, Austin also felt that his work might challenge foundationalist epistemologies, including Kant's, which Frege and Saussure had presupposed.

For both Kant and Frege, the fundamental structures of thought were independent of the structure of language; epistemology had to be able to explain how knowledge of the truths of mathematics and logic was possible. In Frege's work, language encodes the senses of words and sentences, which are ultimately timeless entities, independent of language. Although Frege's sense-reference distinction is a crucial component of Austin's "locutionary act," by the end of *How to Do Things with Words,* a collection of lectures, Austin suggests that even that distinction will have to be rethought. The implication seems to be that looking at the uses of language will lead to a reconsideration of the relations between language and thought. Instead of language encoding thought, thought may ultimately depend on language; reversing traditional priorities, epistemology would presuppose the philosophy of language.

Austin

John Austin developed his distinctive style of linguistic analysis during the postwar period when logical positivism dominated philosophical discussions. The intellectual heritage of British positivists such as A. J. Ayer can be traced back to the Vienna School of the twenties and thirties, overlapping with Ludwig Wittgenstein and Rudolf Carnap, and indirectly with Gottlob Frege. Building on Frege and Russell's work in symbolic logic and the logical analysis of sentences, Wittgenstein in his *Tractatus Logico-Philosophicus* had tried to differentiate logically meaningful statements from all others; for him, "all philosophy is 'Critique of Language.' " Philosophical analysis became the logical analysis of language.

> Most of the propositions and questions to be found in philosophical works are not false but nonsensical. Consequently we cannot give any answers to questions of this kind, but can only point out that they are nonsensical. Most of the propositions and questions of philosophers arise from our failure to understand the logic of our language. (Wittgenstein 1961, 4.003)

The *Tractatus* had a major influence on the development of the Vienna Circle of logical positivism, which Ayer would join in the early thirties. Although there were many differences among the logical positivists, they shared several assumptions. Legitimate statements could be divided into two categories: being either synthetic a posteriori truths that were empirically verifiable or tautologies such as those of logic and mathematics. Every legitimate, nontautological complex statement was a truth-functional product of simple statements that could in principle be confirmed or denied by empirical observations. From such a foundation of empirically verifiable statements, one could then generate all the truths of science.

Austin's two posthumous works, *Sense and Sensibilia* and *How to Do Things with Words* directly attack not only logical positivism but also any type of "foundationalist" epistemology. Toward the end of *Sense and Sensibilia* Austin reveals that his target is nothing less than what he saw as a fundamental tendency in philosophy.

> The pursuit of the incorrigible is one of the most venerable bugbears in the history of philosophy. It is rampant all over ancient philosophy, most

> conspicuously in Plato, was powerfully re-animated by Descartes, and bequeathed by him to a long line of successors. (Austin 1962b, 104)

Logical positivism revealed this tendency in its attempts to locate some set of statements about basic, indubitable data from which the rest of scientific knowledge could be derived. In *Sense and Sensibilia,* Austin shows that Ayer's candidates for such foundational knowledge, "observation sentences" about "sense data," are in fact unattainable; "there isn't, there couldn't be, any kind of sentence which as such is incapable, once uttered, of being subsequently amended or retracted" (Austin 1962b, 112). Austin thought that the search for such a set of sentences was misguided because what makes a sentence appear to be indubitable "is not a matter of what *kind of sentence* I use in making my statement, but of what *the circumstances are* in which I make it" (Austin 1962b, 114). The idea of verifying statements about material objects arises "through the pervasive error of neglecting *the circumstances in which* things are said — of supposing that *the words alone* can be discussed, in a quite general way" (Austin 1962b, 118). Austin concludes with the following flourish:

> For even if we were to make the very risky and gratuitous assumption that what some particular person knows at some particular place and time could systematically be sorted into an arrangement of foundations and super-structure, it would be a mistake in principle to suppose that the same thing could be done for knowledge *in general.* And this is because there *could* be no *general* answer to the questions what is evidence for what, what is certain, what is doubtful, what needs or does not need evidence, can or can't be verified. If the Theory of Knowledge consists in finding grounds for such an answer, there is no such thing. (Austin 1962b, 124)

These considerations are also present in Austin's *How to Do Things with Words,* which opens with a discussion of sentences that look grammatically like statements but according to Austin were not. Austin returns to a point made in his criticisms of logical positivism. Since Kant, philosophers have shown that many statements are "strictly nonsense"; from an inspection of some of these cases, utterances that look like statements turn out to have other functions, and many traditional philosophical perplexities have arisen through "the mistake of taking as straightforward statements of fact utterances which are *either* (in interesting non-grammatical ways) nonsensical *or else* intended as something quite different" (Austin 1962a, 3).

In *How to Do Things with Words,* Austin's target is not logical positivism and the status of sentences about "material things" but linguistic analysis itself. Austin argues that philosophers have taken statements as the primary object of their epistemological analyses; statements are the "foundational" speech acts from which others are derived or at least differentiated. He isolates a set of utterances that grammatically look like statements and are free of verbal peculiarities such as modals ('ought', 'can') or evaluative expressions. These constructions, in the first-person present indicative, are neither true/false nor descriptive/reportorial; instead they constitute "the doing" of an action. He gives four examples:

(1) uttering, "I do" in the course of a marriage ceremony;
(2) uttering, "I name this ship the *Queen Elizabeth*" when smashing a bottle against the stem in a christening ceremony;
(3) "I give and bequeath my watch to my brother" in a will; and
(4) uttering, "I bet you sixpence it will rain tomorrow."

Austin calls these utterances "performatives" as distinguished from "constatives," which are true or false and include statements, descriptions, assertions, and reports. Performatives work by linking the uttering of certain words to specific circumstances of utterance; in the absence of these circumstances (e.g., if a will is not legally prepared in the third example above), performatives can go wrong, or, as Austin puts it, be "unhappy." Austin gives six felicity conditions whose violation will lead to "infelicities." Violating any of the first four conditions can result in misfires in which the desired effect is not achieved. The last two conditions constitute abuses in which one carries out the act but it is not completed (i.e., a promise is made but not acted upon).

(1) There must exist an accepted conventional procedure having a certain conventional effect, that procedure to include the uttering of certain words by certain persons in certain circumstances, and further,
(2) the particular persons and circumstances in a given case must be appropriate for the invocation of the particular procedure invoked.
(3) The procedure must be executed by all participants both correctly and
(4) completely.
(5) Where, as often, the procedure is designed for use by persons having certain thoughts or feelings, or for the inauguration of certain conse-

quential conduct on the part of any participant, then a person participating in and so invoking the procedure must in fact have those thoughts or feelings, and the participants must intend so to conduct themselves, and further

(6) must actually so conduct themselves subsequently. (Austin 1962a, 14–15)

Just as constatives depend on truth conditions for their interpretability, performatives rely on these context-specific "felicity" conditions for their effectiveness. Austin sees the notion of felicity as applying to all conventional acts and even statements. However, in a famous passage that Derrida will pick up on later, he states that

> a performative utterance will, for example, be *in a peculiar way* hollow or void if said by an actor on the stage, or if introduced in a poem, or spoken in a soliloquy. This applies in a similar manner to any and every utterance — a sea-change in special circumstances. Language in such circumstances is in special ways — intelligibly — used not seriously, but in ways *parasitic* upon its normal use. (Austin 1962a, 22)

Austin's specification of these felicity conditions shows that performatives can be systematically analyzed, and therefore overturns the logical positivists' insistence on verifiability as the main criterion for meaningfulness.

In the next several lectures of *How to Do Things with Words,* Austin tries to find some set of features that will uniquely differentiate performatives from constatives. He runs through several possibilities, including the ways in which "happy" performative uses depend on certain statements being true (it must be true that the will is in good order for the performative "I bequeath . . ." to be effective), grammatical criteria (first-person, present indicative, nonprogressive, etc.), and even the ability to take an utterance and put it into the form of an explicit performative ("I will go tomorrow" can be rephrased as "I promise that I will go tomorrow," which removes all ambiguity about what action is performed); none can consistently delimit performatives from constatives, so Austin suggests that "it is time to refine upon the circumstances of 'issuing an utterance.' " Saying anything involves at least:

(1) the utterance of certain noises — a phonetic act;
(2) the utterance of certain "vocables" belonging to the vocabulary, grammar, and stylistic patterns of a given language — a "phatic" act;

(3) the performance of a speech act "with a certain more or less definite 'sense' and a more or less definite 'reference' (which together are equivalent to 'meaning')" — a "rhetic" act.

These three acts make up what Austin calls a "locutionary act," which he then distinguishes from illocutionary and perlocutionary speech acts. The identification of the rhetic component of a locutionary act links the latter to Frege's account of sense and reference. The locutionary act is that which is reported in indirect speech, or *oratio obliqua,* such as "She said that she would be there the next day" as a report of the utterance "I shall be there tomorrow."

Illocutionary acts constitute the ways in which we use locutions to do things such as asking or answering a question, pronouncing a sentence, making promises, making an identification, or giving a description; they are what we do *in* the act of saying something as opposed to the act *of* saying something. A perlocutionary act is what we do *by* saying something, the effect the performance of the speech act has upon the speaker, hearer, audience, or other persons. Austin gives the following examples:

Locution: He said to me, "You can't do that."
Illocution: He protested against my doing it.
Perlocution: He pulled me up, checked me.

Austin focuses on describing and analyzing illocutionary acts, which have been overlooked in his view because philosophers tend to focus on locutionary and perlocutionary acts. Although all could fall under the larger rubric of "uses of language," Austin distinguishes the locutionary from the illocutionary via the "sense and reference" structure of the former and the performativity of the latter, and both of these from the perlocutionary by their conventionality. Illocutionary acts are acts "done as conforming to a convention" (Austin 1962a, 105). Austin then goes through a series of possible tests for the identification of illocutionary forces; one test that survives the change from the performative-constative to the locutionary-illocutionary-perlocutionary classification is the explicit performative. Explicit performatives name the speech event they enact: using the verb "to promise" in the expression "I promise to do X" can, under the right conditions, constitute the making of a promise. Such explicit performatives seem to refer to different types of illocutionary forces; their felicity conditions also specify the nature of their conventionality — that is, what contextual conditions have to hold for them to be "happy."

The earlier constative-performative distinction becomes a subset of the locutionary-illocutionary schema. Constatives are created by concentrating on the locutionary dimension of utterances and abstracting from the illocutionary, "aim(ing) at the ideal of what would be right to say in all circumstances for any purpose, to any audience, &c" (Austin 1962a, 145). Performatives are created by focusing on the illocutionary force of utterances and abstracting from their correspondence with facts. Statements and descriptions have no unique status among many other illocutionary acts, because "truth and falsity are (except by an artificial abstraction which is always possible and legitimate for certain purposes) not names for relations, qualities, or what not, but for a dimension of assessment — how the words stand in respect of satisfactoriness to the facts, events, situations, &c., to which they refer" (Austin 1962a, 148). Instead of a hard and fast dichotomy, we have "more general families of related and overlapping speech acts" that need to be classified and analyzed.

Austin saw that if his line of thinking about speech acts was correct, the implications were dramatic. Not only does truth/falsity become one dimension of linguistic assessment among others, but the fact-value contrast will have to be eliminated and the whole theory of "meaning" based on sense and reference reformulated. Rethinking how language works means decentering the primacy of truth conditions in linguistic analysis; but if linguistic analysis is at the heart of epistemological questions, then such questions will also have to be reconsidered.

What Austin proposes is a radical revision of the priorities of philosophical analysis. In *Sense and Sensibilia* he attacks the logical positivists' analyses of language, and the role played by reality and truth functionality in those analyses, and also hints at the inadequacy of most philosophical theories of knowledge that rely on a decontextualized conception of language. In *How to Do Things with Words* Austin takes aim at the whole model of linguistic analysis that privileges logic and truth functionality, thereby targeting not only Ayer and his followers but also Russell, the early Wittgenstein, and Frege. All these thinkers assume that a perfect language will make clear the relations between language and reference and will be able to precisely express the truths of mathematics, logic, and science. The model presupposes a clear-cut separation between linguistic expressions and what they refer to, between language and some independent reality, whether the latter be a state of affairs, a fact, or a sensory event. Austin focuses on what utterances do; language changes and creates reality. The common notion of a "correspondence" between language

and the extralinguistic is just a certain way of viewing language use. Performatives make explicit the contrast between reference and performance: they look like referring expressions but they create the very event they describe.

In *How to Do Things with Words,* Austin indicates that the line of linguistic analysis he is both trying to extend and criticizing begins with Kant and extends through Frege, the logical atomists, and logical positivists. Kant's distinction between analytic and synthetic truths relied on his understanding of syllogistic logic, which dealt primarily with *inter*propositional relationships. Frege's discovery of quantification theory made it possible for philosophers to systematically analyze *intra*propositional structure, to see how a given proposition consisted of different levels of generality. Austin recognized that Frege's methods had introduced "a revolution in philosophy" that was conceivably "the greatest and most salutary in [that discipline's] history" (Austin 1962a, 3), and one in which he plainly identified his own work as a "piecemeal" beginning.

Austin's strategy in *How to Do Things with Words* is to start with constative uses of speech in which there is a clear distinction between language and reality, and to gradually introduce the role played by the circumstances of utterance in determining the force of a speech act. In so doing, Austin reverses the priority held by language as truth and correspondence over language as action and creation. By the time he formulates his notion of a locutionary act, the Fregean sense-reference distinction is still present as a dimension of assessment, but it is no longer primary; Austin asserts "that the theory of 'meaning' as equivalent to 'sense and reference' will certainly require some weeding-out and reformulating in terms of the distinction between locutionary and illocutionary acts" (Austin 1962a, 148).

Although Austin was not, like Russell, a technical logician, he was amply aware of the magnitude and influence of Frege's discoveries on the foundations of mathematics and logic, as his translation of the *Grundlagen* shows. Yet by the end of *How to Do Things with Words,* Austin seems on the verge of overturning some of Frege's most cherished doctrines regarding the roles of logic, truth, and correspondence in analyzing language. Frege's work was directed at securing the logical foundations of mathematics. As such, it seemed to intersect with that portion of linguistic structure which codified reference and left other portions unanalyzed. Austin's discovery of performativity seemed to point to where the logical analysis of language ended and its contextual analysis began; if performatives did not refer and still could be systematically analyzed, the boundaries of philosophical analysis would be extended beyond the consider-

ations of truth functionality. Clarifying the structure of illocutionary acts would lead to a revision of Frege's notions of sense and reference that, in turn, might lead to rethinking whether reference itself was a logical notion or a speech act.

Frege's account of sense and reference was crucial to Austin's enterprise because it provided the logical tools Austin needed to determine how the semantic meaning of a sentence or proposition is constructed from its component parts. As translator of Frege's *Grundlagen* and as critic of the logical positivist and logical atomist traditions in Oxbridge philosophy, Austin understood what Frege had done and how it had inspired Wittgenstein, Russell, and a generation of British philosophers. From a logical point of view, Frege had given an account of how language, reference, and truth might work together. It was these core insights that Austin preserved in his account of the locutionary act. Illocutionary acts, however, pointed to the contextual dimensions of language use that Frege and his followers had mostly ignored. At the end of *How to Do Things with Words,* Austin's planned trajectory seemed to reverse Frege's. Instead of starting with truth and correspondence, he would focus on the uses of language; specify the illocutionary dimensions of speech acts, including statements and assertions; and eventually revise his account of locutions in the light of his illocutionary investigations. If Austin was right, these revisions would amount to a whole-sale reformulation, from Frege back to at least Kant, of what philosophical analysis consisted in.

Frege

Besides being one of the founders of mathematical logic, Frege is also a key figure in the modern analytic philosophy of language. He directly influenced Wittgenstein and Russell, and several of Austin's basic themes come from Frege. His treatment of the existential quantifiers and his function-argument analysis of propositions have allowed philosophers to see how the logical structure of a proposition is constructed from its component parts, and how its conceptual structure might determine reference. These insights are encapsulated in Frege's famous context principle, which states that the sense of an expression is determined by its role in completing the sense of a proposition or thought. Words and other expressions are not simply names for collections of objects but participate in constructing the abstract meanings of propositions, whose different levels of generality could be logically specified through the judicious manipulation of a proposition's quantifiers; he thereby also laid the

foundation for the semantic analysis of language, in which the referential structure of sentences is seen as a truth-functional product of that structure's parts.

Until Frege's discoveries, the syllogistic logics of his time dealt with interpropositional relations and were unable to deal systematically with multiple levels of generality within a single proposition; a sentence such as 'everybody loves someone who loves everybody' was unanalyzable. Frege's quantification theory solved this problem by analyzing how a proposition is made up of different levels of generality, including proper names and singular terms, the quantifiers 'all' and 'some,' variables, and predicates/functions, such as '_____ loves _____' in the previous example. The meaning of words and other linguistic categories is determined by how each one contributes to the proposition's structure. Building on his discovery of quantification theory, Frege developed a distinction between sense and reference that would prove crucial both to Austin's speech act theory and to the logical analysis of epistemic and modal contexts; without his pioneering work, there would be no contemporary analytic philosophy of language.

The breakthrough work was Frege's *Begriffschrift,* first published in 1879, in which the German philosopher presented an "ideography" that he hoped would be adequate to present proofs and guarantee the validity of arguments. Symbolic language was not a model of the actual chain of reasoning that led to a conclusion but, rather, an attempt to lay out in the clearest manner possible all the logical presuppositions and entailments necessary to ensure the validity of argumentation. In order to accomplish this goal, one needed a symbolic language within which any statement of mathematical theory could be framed (a formalized language) and effective criteria for recognizing whether any particular collocation of symbols was a formula belonging to that language (rules for well-formedness). Given the adequacy of such a language, Frege then saw the need for formal rules of proof—that is, rules for determining which sequences of the language were valid.

Frege found natural language inadequate for this task and thus developed the "formal language for pure thought" of the *Begriffschrift,* whose "first purpose . . . is to provide us with the most reliable test of the validity of a chain of inferences and to point out every presupposition that tries to sneak in unnoticed, so that its origin can be investigated" (Frege 1970a, 6). He characterizes his formal language as resembling a microscope, whose resolving power is far greater than the natural eye of language. Natural language, because of the range of functions it serves, cannot be seen as precise enough for the de-

tailed work of analyzing the logical morphology of human thought. Yet Frege's claims for his ideography were not as modest as they may at first seem, for Frege held that his system provided a secure foundation for philosophy itself.

This immodesty, however, was not unfounded. As J. Van Heijenoort points out in his brief introduction to the English translation of the *Begriffschrift,* Frege's work lays out "the truth-functional propositional calculus, the analysis of the proposition into function and argument(s) instead of subject and predicate, quantification theory, a system of logic in which derivations are carried out exclusively according to the form of the expressions, and a logical definition of the notion of mathematical sequence" (Frege 1970a, 1), along with assorted ideas about how to "regiment" natural language into propositional form. Given the nature of these discoveries, it is not surprising that they form the basis for Frege's later work, both on language and thought and in mathematical logic.

In the *Begriffschrift* Frege uses a mathematical analogy to distinguish two kinds of symbols. The first signs are marks of generality resembling mathematical symbols that designate an indeterminate number or function, such as the letters in the formula (a + b)c = ac + bc. The other category of signs contains marks of particularity, such as '+', '−', '0', '1', or '2'. Frege then introduces an assertion or judgment sign, '⊢'. What follows this sign is the content of the judgment. If the vertical stroke is omitted, the remaining horizontal line indicates that the following sign (or signs) is being considered as a "mere combination of ideas" without regard to its potential truth value. If '⊢A' stands for "the judgment that opposite magnetic poles attract each other," then '— A' serves merely to produce the idea of the mutual attraction of opposite magnetic poles. The sign '—' becomes paraphrastically equivalent to "the proposition that." Not everything can provide content for a judgment. Frege argues that the idea "house" cannot, because by itself it does not constitute a proposition capable of a truth-functional determination.

Frege then proceeds to show that the subject-predicate division plays no role in his account of logical form. The conceptual contents of two judgments are the same if the logical consequences derivable from the first are equivalent to those derivable from the second. In ordinary language, the subject position is important because it is "where we put that to which we wish especially to direct the attention of the listener" (Frege 1970a, 12). Such distinctions are not relevant to Frege's formal language, however, because they do not influence the possible logical entailments.

> Now, all those peculiarities of ordinary language that result only from the interaction of speaker and listener — as when, for example, the speaker takes the expectations of the listener into account and seeks to put them on the right track even before the complete sentence is enunciated — have nothing that answers to them in my formula language, since in a judgment I consider only that which influences its *possible* consequences. Everything necessary for a correct inference is expressed in full, but what is not necessary is generally not indicated; *nothing is left to guess work.* (Frege 1970, 12)

In modern terminology, pragmatic discourse conditions are irrelevant. If one insists on a subject-predicate notion, Frege says one should consider the whole content of the judgment to be the subject; the judgment stroke '⊢' would be a predicate equivalent to " — is a fact."

After describing his theory of judgment, Frege introduces an expression that is equivalent to the conditional 'B → A'. Frege describes this as the judgment that the case of B being affirmed and A denied does not hold while one of the other three possibilities (B affirmed, A affirmed; B denied, A affirmed; B denied, A denied) does hold. He then introduces his only rule of inference, *modus ponens:* From 'B → A' and 'B', one can validly infer 'A'. Frege recognizes that there are other valid forms of inference but retains only one for his own system for reasons of perspicuity and simplicity. Frege then introduces a negation sign, which in his ideography expresses the circumstance that the content does not hold. For example '−A' means 'A does not occur'. He then proceeds to cases where negation and the conditional are combined. His example is the situation in which 'it is not the case that B is affirmed and A is not denied,' which in turn means 'it is not the case that both A and B are affirmed' ('B → −A' in modern symbolic notation). Frege then shows how to define exclusive and inclusive 'or' in terms of his two operators, and points out that 'and' is describable in terms of the conditional and negation or that the conditional is definable in terms of 'and' and negation. He also points out that 'and' (*und*) and 'but' (*aber*) are conceptually equivalent but differ in nonreferential meaning because 'but' hints that the following proposition is unexpected.

Frege then turns to the problem of statements of identity. Identity equations are essential because they are the foundation for definitions. They differ from statements of conditionality and negation in that they apply to names and not to contents. Whereas in other contexts signs stand for their contents and sign

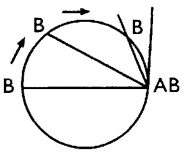

Figure 1.1 Fregean Circles

combinations express nothing but relations among the contents represented by the signs, in the context of an identity statement the names are also being talked about—they "display their own selves"—and serve to indicate that two (or more) names have the same content. Frege asks us to assume that there is a fixed point A on the circumference of a circle and to let a straight line rotate around this point A (fig. 1.1). When that line forms a diameter of the circle, let the point opposite A where the diameter crosses the circumference be labeled 'B'. Then let the point of intersection between the line and the circumference continue to be called B as the line is rotated, with the location of point B varying accordingly. Thus the location of point B is indeterminate until the corresponding line is defined. It can then be asked, What point corresponds to the position of the line when the latter is perpendicular to the diameter AB? In that situation, the specified point B is equivalent to point A. The names 'A' and 'B' thus have the same content, and yet it would have been impossible to use one name from the beginning, as the whole exercise shows that this point can be specified in two different ways, by direct stipulation or by specifying that point B is a straight line perpendicular to the circle's diameter.

Frege's basic point is that the same content can be specified in different ways. One can thus construct a judgment in which there are two ways of determining the same content, and the two distinct names used correspond to these two modes of determination. Since the identity statement asserts that sign A and sign B have the same conceptual content, the two signs can be substituted for one another without producing any effect at the conceptual level; therefore, such substitutions will not affect any valid arguments. Frege will later modify his analysis of identity in his 1891 article "On Sense and Reference" (Frege 1970c) by introducing the notion of sense, but that notion will not change his account of the extensional aspects of quantification theory.

Frege then introduces his idea of function, which replaces the old subject-

predicate distinction he had criticized earlier. In an expression of mathematical function, such as $x^3 + 2x^2 + 4$, one distinguishes between the argument positions indicated by the variables and the "incomplete," or unsaturated, relation $(\ )^3 + 2(\ )^2 + 4$. Frege applied the same model to language.

> If in an expression, whose content need not be capable of becoming a judgment, a simple or compound sign has one or more occurrences and if we regard that sign as replaceable in all or some of these occurrences by something else (but everywhere by the same thing), then we call the part that remains invariant in the expression a function, and the replaceable part the argument of the function. (1970a, 22)

In his later work, Frege will expand his notion of function and also make a distinction between function expressions and what they designate. Frege's particular example is the proposition that hydrogen is lighter than carbon dioxide as it is expressed in his formal language. The sign for hydrogen can be replaced by the sign for oxygen or nitrogen, so that "oxygen" or "nitrogen" enters into relations in which "hydrogen" previously contracted. If an expression can be so altered, there is a stable component represented by the totality of relations, which he calls the function, and a replaceable portion, which is the argument(s). In the example, "hydrogen" is the argument and "being lighter than carbon dioxide" the function, and "oxygen" also becomes an argument of the same function. If one treats "carbon dioxide" as the argument, then "being heavier than hydrogen" is the corresponding function. The line of reasoning can be extended to functions of more than two arguments.

> If, given a function, we think of a sign that was hitherto regarded as not replaceable as being replaceable at some or all of its occurrences, then by adopting this conception we obtain a function that has a new argument in addition to those it had before. (1970a, 23)

So "the case that hydrogen is lighter than carbon dioxide" can also be regarded as a function of two arguments, "hydrogen" and "carbon dioxide."

Frege also warns that not everything that occurs in the subject position can be an argument, that normal linguistic usage can be deceptive as regards a proposition's true logical form. If the two propositions "The number 20 can be represented as the sum of four squares" and "Every positive integer can be represented as the sum of four squares" are compared, it seems that "being representable as the sum of four squares" is a function that can take the arguments

"the number 20" and "every positive integer." However, these two expressions are not of the same level of generality; "every positive integer" is not an independent idea but is instead reliant on the context of the sentence for its meaning. Frege will later call such expressions incomplete. Frege represents an indeterminate function of the argument A as 'F(A)'; 'F(A, B)' is a complex sign that is a function of the two arguments A and B taken in that order. Thus, F (A, B) ≠ F(B, A). The formula 'F(A)' is to be read as the judgment that 'A has the property F', and 'F(A, B)' as 'B stands in relation F to A'.

Frege then introduces his treatment of quantification. His function-argument ideography, when combined with his treatment of the quantifiers and the concomitant notation of bound variables, allowed him to clearly express multiple levels of generality within a given judgment. His first move was to define the universal quantifier, '(x)(Fx)'. This expression states that the function in question, 'Fx', is a fact no matter what is assigned to its argument place; that is, 'for any object, call it a, it has the property F', or 'everything is F'. His ideography also specifies the scope within which the sign of generality (in modern parlance, a "variable") remains valid. The sign has a fixed meaning only within its own scope; thus within one judgment the same variable can have different meanings attributed to it as long as the scopes in each instance do not overlap. Within one judgment the same variable can occur in different scopes without the meaning attributed to the variable in one scope being extended to that variable in any other. The scopes of one variable can include another, in which case the variables must be different, although *within* a given scope any variable can be replaced by any other variable sign as long as the replacements are uniform. Here is Frege's example, transcribed in modern notation: '(x)(y)(Bxy → Aw)' but not '(w)(w)(Bww → Aw)', where the replacement has not been systematic (i.e., each different letter has not been replaced by a different letter in all substitution places). Frege has thus provided a notation for binding variables and indicating substitution rules that will preserve the validity of arguments. He then combines these insights with his negation operator, thus generating equivalents to the modern sign '∃x(Fx)' (there is some x with property F). In his notation, the expression '∃x(Fx)' is equivalent to 'not everything is F' (which is equivalent to '−(x)(Fx)'), or 'something is not F'. In order to assert that something has property F, the necessary sign is '−(x)(−Fx)' (which is equivalent to '−x−(Fx)', 'something is not not F' or 'something is F'. Frege also allows one universal quantifier to be placed within the scope of another in such a way as to produce embedded levels of generality, a type of complexity

not even considered by earlier logicians even though ordinary and scientific language equivalents can be found for expressions containing multiple levels of generality. For example, 'everyone loves some lover' can be represented in one of its readings as '$(x)(\exists y)(Cx.By \rightarrow Dxy)$', where 'C' is '_____ is a person', 'B' is '_____ is a lover', and 'D' stands for '_____ loves _____'. Finally, Frege produces the square of logical opposition for his quantifiers and a logical definition of mathematical sequence.

Frege was well aware of the magnitude of his discoveries in the *Begriffschrift*. In one fell swoop, he had solved a long list of logical problems that had befuddled philosophers since the time of Aristotle. These discoveries were to influence all of his subsequent work. The quantificational theory that Frege had invented allowed him to see sentences as being constructed in an ordered sequence of steps. Insofar as sentences express different levels of generality, then Frege's logical apparatus allowed him to show clearly how the different orders of generality were combined to form sentences that had a determinate truth value. The analogy was taken directly from mathematics. For example, if one has an expression such as '$(2 + 4) \times 3 = 18$', the derivation of a particular value has to occur in a sequence of steps in which the addition sign functions as an operator and is clearly different from the numerals. In the first stage of addition, the parentheses serve to indicate that the whole expression '$2 + 4$' is to be taken as a unit. A second stage consists of combining '$2 + 4$' with the number '3' by means of the multiplication sign. Frege's particular accomplishment was to see that sentences could be treated in the same way. Different levels of generality were to be seen as combining in a hierarchical collocation.

Frege was also the first logician to see clearly the relationships between problems of identity and those of reference. In the *Begriffschrift* Frege points out that the terms appearing in a true statement of identity may be substituted for each other in any true statement and the resulting statement will be true. One can interchange proper names or singular terms that denote the same object in all contexts without affecting the truth value of the expressions of which they are a part.

Linsky (1977, 115–17) has usefully summarized the ways in which identity, quantification, and singular reference work together. The principle of substitutivity underlies both the concept of singular reference and that of quantification, and it is an integral part of a coherent semantics for quantification theory. In logic, in order to evaluate sentences from a truth-functional point of view, we start from an open sentence such as 'Fx', which is not evaluable; it

becomes evaluable when its free variable, in this case 'x', is bound by a quantifier or replaced by a singular term or proper name. Thus, from the open sentence 'Fx' we can construct the closed sentences '$\exists$x(Fx)' and 'Fa', where 'a' is a singular term or proper name.

Frege's account of substitutivity and its relation to singular reference can be understood in terms of four points. First, if we replace a singular term in all of its occurrences in a sentence by an appropriate variable, we construct a paradigmatic open sentence. Second, if the sentence in which the singular term occurred is true, then that singular term refers to an object that satisfies the open sentence constructed by substituting a free variable for the singular term. Third, an object satisfies such an open sentence only if replacing the open sentence's free variable by any singular term making reference to that object turns the open sentence into a true statement. Fourth, the replacement of a singular term in a true statement by any other singular term referring to the same object, leaves the truth value of the original statement unchanged. That is, terms of a true identity statement refer to the same thing. Substitutivity and singular reference are thus linked. Any failure in the substitutivity of coreferential singular terms entails a failure of reference.

Frege's treatment of quantification theory makes clear the relationship between the principle of substitutivity and quantification. The sentence 'There is an x such that F of x' ('$\exists$x(Fx)') is true if and only if there is some object (at least one) in the range of the variable 'x' that satisfies the open sentence 'F of x' ('Fx'). If the sentence 'F of a' ('Fa') is true where 'a' is a singular term and satisfies the principle that the intersubstitution of coreferential terms preserves truth value, then 'a' denotes the object satisfying F of x. If the expression 'There is an x such that F of x' ('$\exists$x(Fx)') is true, then what 'There is an x such that F of x' signifies is merely that some object has the property F and that the particular mode of presentation of the object of the variable 'x' is not at issue. We have abstracted from 'F of a' ('Fa'), which specifies a particular mode of presentation of the object designated by 'a', to the sentence 'There is an x such that F of x', in which the particular mode of presentation of the object by 'a' is irrelevant. Only if 'a' fails to refer in the expression 'Fa' is the inference '$\exists$x(Fx)' that we draw from 'Fa' invalid. A classic example of such failure is the sentence 'Pegasus does not exist', where the inference '$\exists$x(Fx)' ('There is an x such that x does not exist') is invalid because nothing can satisfy the open sentence 'x does not exist'. When the principle of substitutivity and singular reference is combined with quantification, any failure of substitutivity entails a failure of

reference and, accordingly, an existential failure for the term that fails substitutivity. This is because the inference whereby we pass from 'Fa' to '$\exists$x(Fx)' is valid regardless of the mode of specification associated with the name 'a'. The expression '$\exists$x(Fx)' is true, as long as some object, no matter how designated, has the property F. Any other name that designated the same object would also support the inference '$\exists$x(Fx)'. It is thus impossible to quantify over singular terms that fail substitutivity. Contexts in which the principle of the intersubstitutability of coreferential singular terms fails are called, in modern language, "referentially opaque," and quantification into referentially opaque contexts is impossible. Frege's work establishes that the principle of substitutivity is an integral part of the semantics of classical quantification theory.

In order to be able to carry out the project he had outlined in the *Begriffschrift*, Frege had to develop an analysis of natural language using those aspects of natural language which would be analyzable as forming the content of judgments. Since judgments would form the basis of the relationships described in logical proofs, he needed to analyze the meaning of such statements. Frege thus found himself forced into performing a semantic analysis of natural language, or at least those parts of natural language which could be regimented for the purposes of logical manipulation. In the *Begriffschrift* he had made some preliminary analyses of sentence structure, mainly to point out where his logical notation diverged from normal discourse practices. For example, certain regular differences in meaning, between active and passive sentences, for instance, or between the words 'and' and 'but', were not part of the content of judgments.

In his classic paper "On Sense and Reference" in 1891, Frege applies these insights to natural language, and to the problem of proper names in particular. This work is an extension of his logical ideas, abstract function theory foremost among them, to language. He begins with a question about equality. Such an issue, of course, immediately touches on issues of reference and quantification, and his introduction of the notion of "sense" will tie these issues to problems in epistemology. Frege asks whether equality is a relation between objects or between names of objects. In the *Begriffschrift* he had said that equality had to be a relation between the names of objects, because 'a = a' and 'a = b' are statements of differing cognitive value. While the statement 'a = a' is a priori true, it does not extend our knowledge, whereas 'a = b' can. If equality were merely a relationship between the entities that 'a' and 'b' designate or refer to, and if 'a = b' is true, then 'a = b' would not differ from 'a = a'. It would seem,

then, that 'a' and 'b' designate the same thing, so the signs themselves appear to be under discussion. The interpretation provided by Frege in the *Begriffschrift* is that such a statement of identity means that 'a' and 'b' are names that designate the same object.

In "On Sense and Reference," Frege states that this earlier interpretation of identity, although logically adequate, seems counterintuitive from an epistemological standpoint. It could not be that such identity statements are talking only about the signs themselves and that identity is no more than a relation between signs. If we make the statement that 'a = b', we could not be saying simply that 'a' and 'b' are names for the same object, because while the choice of a given name for an object is arbitrary, identity statements do seem informative. We seem to learn something about the world when we discover such identities, not merely something about names. The identity statement seems to have cognitive value only if the difference between the signs themselves corresponds to a difference in the mode of presentation of that which is designated. We need to distinguish between a sign's denotation (what object it refers to) and its mode of presentation, or sense. In Frege's 1879 *Begriffschrift* example of the two points connected by a line on the circumference of a circle, we can already discern an implicit concept of two different modes of presentation. In his actual definition of identity, however, Frege does not address the mode of presentation. He does not need to because in standard quantification theory, the mode of presentation of two names that designate the same object is irrelevant to the truth value of the sentences that contain them. For epistemology, however, it is clear that the mode of presentation is important.

In his later work, Frege provides the example of a triangle subdivided by three lines that connect the vertices of the triangle with the midpoints of the opposite sides. If 'a', 'b', and 'c' designate these lines, then the point of intersection of lines a and b is the same as the point of intersection of lines b and c. There are three different designations for the same point, and the names — for instance, 'point of intersection of a and b', 'point of intersection of b and c' — also indicate the mode of presentation; thus the statement of identity contains cognitive value, or real knowledge. The names of these points of intersection all have the same referent but differ in their mode of presenting that referent. Frege calls the sign's mode of presentation its "sense." He also brings up the example of 'the Evening Star' and 'the Morning Star', terms that Frege sees as designating the same object using different modes of presentation. Any singular term refers to an object — its referent — and contains a mode of presenta-

tion—the "sense" of the sign. Frege then makes a distinction between the sense, the referent, and the idea associated with a given sign. The referent of a sign is an object that is perceivable by the senses, while the idea is an internal image arising from a person's sensory impressions and associated psychological states. The idea is something individual, something linked to a particular person. Senses and ideas are not necessarily connected, even for a single individual. The idea is subjective; it can vary from person to person. Frege postulates, for instance, that a painter, a horseman, and a zoologist will probably associate different ideas with the name Bucephalus. Frege also offers the example of a person observing the moon through a telescope. The moon is the referent, the real image in the lens of the telescope is the sense, and the particular person's retinal image of the real image in the lens of the telescope is his or her idea. In most circumstances there is a regular connection between a sign, its sense, and its denotation. Each sign has a definite sense, and each sense has a definite denotation. The denotation of a proper name or singular term is the object itself designated by the term. A proper name expresses its sense and stands for, or designates, its denotation. But in certain circumstances, such as fictional discourse, an expression has a sense but no denotation.

In normal discourse words are used to talk about their denotation. There are special circumstances, however, in which words are not used to talk about their ordinary denotation; in reported speech, for example, the words themselves are the object of discussion. In such circumstances the speaker's words refer to the words of another speaker, and only the latter have their usual denotation. Another example of a denotational switch is provided by indirect discourse and other "oblique" contexts in which we talk about the senses of expressions. In these cases a word does not have its customary denotation but designates its sense; it is used indirectly. The indirect reference in such instances is the customary sense of the expression being discussed. In the sentence 'Frege believes that the Morning Star is Venus', 'the Morning Star' and 'Venus' stand for, or designate, their indirect referents rather than their customary referents. If they denoted their normal referents, then we would be able to substitute any coreferential expression for them without changing the truth value of the sentence. However, substituting 'Evening Star' for 'Morning Star' produces 'Frege believes the Evening Star is Venus', which may be a false statement. If in such oblique contexts expressions refer to their ordinary senses, then the latter sentence becomes something like 'Frege believes that the brightest shining object seen in the heavens right after sunset is Venus', which is different in

cognitive value from the sentence 'Frege believes that the brightest shining object seen in the heavens right before sunrise is Venus'.

Frege then applies the sense-reference distinction to whole sentences. For Frege, each declarative sentence expresses a thought, so he asks whether this thought is to be regarded as the sense or the reference of the sentence. Assuming that a given sentence has a reference, then we should be able to replace one word of the sentence by another word having the same reference but a different sense without changing the reference of the sentence. We find, however, that such substitutions do affect the cognitive value of the sentence; that is, the thought changes. For example, the thought expressed in the sentence 'The Morning Star is a body illuminated by the sun' differs from that of 'The Evening Star is a body illuminated by the sun'. If someone did not know that the Evening Star was the Morning Star, he might hold the former sentence to be true and the latter false. Frege then concludes that the thought cannot be the reference of the sentence but rather must be its sense.

If a thought is the sense of a sentence, then what is the referent of a sentence? From the fact that we concern ourselves with the reference of a part of the sentence, Frege concludes that we generally recognize and expect a reference for the whole sentence itself. When does the notion of the reference of a sentence make a difference? It makes a difference in precisely those cases in which we are concerned with a sentence's truth value. As Frege puts it, "It is the striving for truth that drives us always to advance from the sense to the reference" (1970b, 63). Frege then concludes that the referent of a sentence is its truth value. If sentences have sense (thoughts) and referents (the true or the false), then by extension they must also have indirect referents and senses. If we take a false sentence such as 'the sun revolves around the earth' and embed it in an oblique context such as 'Aristotle believed that . . .', we obtain the sentence 'Aristotle believed that the sun revolved around the earth', which happens to be true. If truth values are objects that sentences refer to, then sentences that refer to the same truth value are intersubstitutable. The principle of unrestricted substitution of coreferential terms does not work for sentences in oblique contexts such as in our example, however, for this would imply that Aristotle believed every false sentence. In oblique contexts, sentences do not have their customary reference (a truth value) but instead designate their ordinary senses or the thoughts that such sentences express. In these cases, we cannot substitute sentences that have similar truth values but may substitute only those which express the same thought or proposition.

What Frege is doing is trying to establish the consistency of his principle of

substitutivity. Substituting one coreferential name for another with the resulting shift in truth value does not contradict this principle, because in opaque contexts names do not have their usual referents and so are not indicating the same referent they usually designate. Instead, they are referring to their senses. Actually, Frege is compelled to take this route. If a name stands for an object, then the whole statement of which it is a part is true or false if and only if the complex predicate formed by omitting that occurrence of the name is true or false for that object. By the principle of existential generalization, a predicate is true or false with regard to a given object, regardless of how the object is presented. Thus filling the argument place of the predicate by any other name also referring to that object must result in a sentence whose truth value is equal to that of the original sentence (the sentence before the omission of the name). An exception means that the name did not have the same referent as the one whose place it took.

Although John Austin is often considered the founder of speech act philosophy, the basic parameters were actually set by Frege in his distinction between the force and sense and reference of expressions. As we saw in the discussion of the *Begriffschrift* above, Frege distinguishes in his ideography between the content of a thought, indicated by the content stroke ' — ', and the judgment that the thought is true, indicated by the assertion stroke '|'. What follows the content stroke is an idea that can be asserted or judged to be true; quantificational analysis uncovers the internal structure of the idea in question. In his later work Frege explicitly connects assertion and judgment, commenting that his judgment stroke is an indication of assertoric force, and that he considered this "dissociation of assertoric force from the predicate" one of his key discoveries (Frege 1979, 184).

> To think is to grasp a thought. Once we have grasped a thought, we can recognize it as true (make a judgment) and give expression to our recognition of its truth (make an assertion). (Frege 1979, 185)

Since a thought is the sense of a sentence, the assertion of a sentence involves understanding its sense, making a judgment of its truth or falsity, and expressing that judgment. Frege further connects assertoric force with "the indicative mood of the sentence that forms the main clause" (1979, 198) and notes that in natural language (fiction excluded), the only place in which we can express thoughts without asserting them is in subordinate clauses such as indirect discourse.

Frege distinguished between force and (propositional) content because he

believed that logic studied the forms of valid inference among judgments or assertions and that it was therefore necessary to identify what assertions are. Natural language tends to obscure the relation between assertion and propositional content in that a sentence in the indicative mood can have assertive force and be subject to truth-functional evaluation when in isolation, while the same sentence used in a subordinate clause or as the antecedent (or consequent) of a conditional will lack assertive force (the assertive force applies to the complex sentence or conditional as a whole) but still contribute to the truth functionality of the whole expression. Since the same linguistic form might in one context have assertoric force and in another not, and since Frege wanted his *Begriffschrift* expressions to be context-independent and unambiguous ("Everything necessary for a correct inference is expressed in full" [Frege 1970a, 12]), Frege invented two signs, the assertion stroke, '|', and the content stroke, '—', to carefully separate what ordinary language seemed to obscure. A judgment consists of two moments, that of thinking of a content and that of judging it to be true; the two signs and their combination, '⊢', thus perspicuously represent these two components and the way they combine to form a judgment.

Frege offers two paraphrases for the content stroke: "the circumstance that" and "the proposition that," followed by some sentence. The judgment content's lack of assertoric force is supposed to be indicated by the noun clause beginning with "that," which by itself is an incomplete expression and could not normally be used to make an assertion ("that two plus two is four"). These paraphrases ultimately depend on the behavior of their noun-clause constituents; these noun-clauses are identical to those expressions which could complete the assertion stroke, namely, "— is a fact" or "it is a fact that. . . ." Frege interpreted these noun clauses as referring to an abstract, Platonic entity, a proposition. All true mathematical statements have the same referent — namely, the true — and true mathematical identities indicate that their constituents have the same referent but different senses. Since Frege believed that mathematical truths existed independently of our discovery of them, it was natural for him to separate the psychological aspects of assertion and thinking from the realm of mathematical truths that were the object of such processes.

When Frege turned to language, he found in the phenomenon of indirect discourse properties analogous to his treatment of sense and reference in mathematics. The subordinate clause of indirect discourse can be an indicative sentence that standing alone would be an assertion but in the embedded context lacks any assertoric force. In such instances, the distinction between a proposi-

tion/thought and its assertion seems to parallel the one between mathematical truths and their assertion. Frege then expanded his analysis to include the relation between thoughts and mental states or activities, since verbs of thinking, believing, and intending could also introduce subordinating constructions similar to those of indirect discourse. In such contexts the referent of the embedded clause is its ordinary sense; the subordinate clause that follows a verb of speaking or thinking would therefore be a thought. What a propositional attitude is about, that is, the object or content of consciousness, is the referent of the sentence expressed in the subordinate clause of indirect discourse.

As we have seen, Austin used the distinctions he articulated as the basis for what he will call constatives and locutionary acts. Since assertion (in German, *bejahen*) is a performative, Frege uncovered in his analysis of content, judgment, and assertion the basic distinction between locution and illocutionary force that is at the heart of Austin's analysis of performativity. What Frege identifies as the content of a judgment is the descriptive or ideational component of constatives and locutions. In addition, since he is able to offer an analysis of the logical syntax of such descriptive content strictly in terms of truth-functional structure, Frege seems to have distinguished those aspects of language and meaning that are independent of any reference to the act of speaking. Frege's work was directed at securing the logical foundations of mathematics. As such, it seemed to focus on those portions of linguistic structure which codified reference and left other portions unanalyzed. By expanding the notion of force beyond Frege's analysis of assertion to include speech acts such as promising and declaring, Austin also seemed to expand the horizons of philosophy. The philosophy of language would have to give up the security of encoding truth in favor of the freedom of language use.