

Theories of meaning

for the *Stanford Encyclopedia of Philosophy*

October 21, 2009

1	Two kinds of ‘theory of meaning’	1
2	Semantic theories	3
	2.1 Propositional semantic theories	4
	2.1.1 The theory of reference	4
	2.1.2 Theories of reference vs. semantic theories	5
	2.1.3 The relationship between content and reference	7
	2.1.4 Character and content, context and circumstance	8
	2.1.5 Possible worlds semantics	12
	2.1.6 Russellian propositions	15
	2.1.7 Fregean propositions	18
	2.2 Non-propositional theories	22
	2.2.1 The Davidsonian program	23
	2.2.2 Chomskyan internalist semantics	25
3	Foundational theories of meaning	26
	3.1 Mentalist theories	27
	3.1.1 The Gricean program	27
	3.1.2 Meaning, belief, and convention	29
	3.1.3 Mental representation-based theories	30
	3.2 Non-mentalist theories	30
	3.2.1 Causal origin	31
	3.2.2 Truth-maximization and the principle of charity	32
	3.2.3 Regularities in use	33
	3.2.4 Social norms	34

1 Two kinds of ‘theory of meaning’

In “General Semantics,” David Lewis wrote

“I distinguish two topics: first, the description of possible languages or grammars as abstract semantic systems whereby symbols are associated with aspects of the world; and, second, the description of the psychological and sociological facts whereby a particular one of

these abstract semantic systems is the one used by a person or population. Only confusion comes of mixing these two topics.” (Lewis (1970), 19)

Lewis was right. Even if philosophers have not consistently kept these two questions separate, there clearly is a distinction between the questions ‘What is the meaning of this or that symbol (for a particular person or group)?’ and ‘In virtue of what facts about that person or group does the symbol have that meaning?’

Corresponding to these two questions are two different sorts of theory of meaning. One sort of theory of meaning — a *semantic theory* — is a specification of the meanings of the words and sentences of some symbol system. Semantic theories thus answer the question, ‘What is the meaning of this or that expression?’ A distinct sort of theory — a *foundational theory of meaning* — tries to explain what about some person or group gives the symbols of their language the meanings that they have. To be sure, the shape of a correct semantic theory may place constraints on the correct foundational theory of meaning, or vice versa; but that does not change the fact that semantic theories and foundational theories are simply different sorts of theories, designed to answer different questions.

To see this distinction, it may help to consider an analogous one. Imagine an anthropologist specializing in table manners sent out to observe a distant tribe. One task the anthropologist clearly might undertake is to simply describe the table manners of that tribe — to describe the different categories into which members of the tribe place actions at the table, and to say which sorts of actions fall into which categories. This would be analogous to the task of the philosopher of language interested in semantics; her job is say what different sorts of meanings expressions of a given language have, and which expressions have which meanings.

But our anthropologist might also become interested in the nature of manners; he might wonder how, in general, one set of rules of table manners comes to be the system of etiquette governing a particular group. Since presumably the fact that a group obeys one system of etiquette rather than another is traceable to something about that group, the anthropologist might put his new question by asking, ‘In virtue of what facts about a person or group does that person or group come to be governed by a particular system of etiquette, rather than another?’ Our anthropologist would then have embarked upon the analogue of the construction of a foundational theory of meaning: he would then be interested, not in which etiquette-related properties particular action types have in a certain group, but rather the question of how action-types can, in any group, come to acquire properties of this sort.¹ Our anthropologist might well be interested in both sorts of questions about table manners; but they are, pretty clearly, different questions. Just so, semantic theories and foundational theories of meaning are, pretty clearly, different sorts of theories.

¹For an illuminating discussion of the analogies between semantic rules and systems of etiquette, see McKeown-Green (2002).

The term ‘theory of meaning’ has, in the recent history of philosophy, been used to stand for both semantic theories and foundational theories of meaning. As this has obvious potential to mislead, in what follows I’ll avoid the term which this article is meant to define and stick instead to the more specific ‘semantic theory’ and ‘foundational theory of meaning.’ ‘Theory of meaning’ *simpliciter* is to be understood as ambiguous between these two interpretations.

Before turning to discussion of these two sorts of theories, it is worth noting that one prominent tradition in the philosophy of language denies that there are facts about the meanings of linguistic expressions. (See, for example, Quine (1960) and Kripke (1982); for critical discussion, see Soames (1999).) If this sort of skepticism about meaning is correct, then there is neither a true semantic theory nor a true foundational theory of meaning to be found, since the relevant sort of facts simply are not around to be described or analyzed. Discussion of these skeptical arguments is beyond the scope of this entry, so in what follows I’ll simply assume that skepticism about meaning is false.

2 Semantic theories

The task of explaining the main approaches to semantic theory in contemporary philosophy of language might seem to face an in-principle stumbling block. Given that no two languages have the same semantics — no two languages are comprised of just the same words, with just the same meanings — it may seem hard to say how we can say anything about different views about semantics in general, as opposed to views about the semantics of this or that language. This problem has a relatively straightforward solution. While it is of course correct that the semantics for English is one thing and the semantics for French something else, most concede that the various natural languages should all have semantic theories of (in a sense to be explained) the same form. The aim of what follows will, accordingly, be to introduce the reader to the main approaches to natural language semantics — the main views about the right form for a semantics for a natural language to take — rather than a detailed examination of the various views about the semantics of some particular expression. (For some of the latter, see NAMES, DESCRIPTIONS, PROPOSITIONAL ATTITUDE REPORTS, and NATURAL KINDS.)

Most philosophers of language these days think that the meaning of an expression is a certain sort of entity, and that the job of semantics is to pair expressions with the entities which are their meanings. For these philosophers, the central question about the right form for a semantic theory concerns the nature of these entities. Because the entity corresponding to a sentence is called a *proposition*, I’ll call these *propositional semantic theories*. However, not all philosophers of language think that the meanings of sentences are propositions, or even believe that there are such things. Accordingly, in what follows, I’ll divide the space of approaches to semantics into propositional and non-propositional semantic theories.

2.1 Propositional semantic theories

The easiest way to understand the various sorts of propositional semantic theories is by beginning with another sort of theory: the theory of reference.

2.1.1 The theory of reference

The theory of reference is a theory which, like a propositional semantic theory, pairs the expressions of a language with certain values. However, unlike a semantic theory, a theory of reference does not pair expressions with their meanings; rather, it pairs expressions with the contribution those expressions make to the determination of the truth-values of simple sentences in which they occur. (It will become clear later why the restriction to ‘simple sentences’ is required.)

This construal of the theory of reference is traceable to GOTTLOB FREGE’s attempt to formulate a logic sufficient for the formalization of mathematical inferences (see especially Frege (1879/1997); see also Frege (1892/1960).) The construction of a theory of reference of this kind is best illustrated by beginning with the example of proper names. Consider the following sentences:

(1) Barack Obama is the 44th president of the United States.

(2) John McCain is the 44th president of the United States.

(1) is true, and (2) is false. Obviously, this difference in truth-value is traceable to some difference between the expressions ‘Barack Obama’ and ‘John McCain.’ What about these expressions explains the difference in truth-value between these sentences? It is very plausible that it is the fact that ‘Barack Obama’ stands for the man who is in fact the 44th president of the United States, whereas ‘John McCain’ stands for a man who is not. This indicates that the reference of a proper name — its contribution to the determination of truth conditions of simple sentences in which it occurs — is the object for which that name stands.

Given this starting point, it is a short step to some conclusions about the reference of other sorts of expressions. Consider the following pair of sentences:

(3) Barack Obama is a Democrat.

(4) Barack Obama is a Republican.

Again, the first of these is true, whereas the second is false. We already know that the reference of ‘Barack Obama’ is the man for which the name stands; so, given that reference is power to affect truth-value, we know that the reference of *predicates* like ‘is a Democrat’ and ‘is a Republican’ must be something which

combines with an object to yield a truth-value. Accordingly, it is natural to think of the reference of predicates of this sort as functions from objects to truth-values. The reference of ‘is a Democrat’ is that function which returns the truth-value ‘true’ when given as input an object which is a member of the Democratic party (and the truth-value ‘false’ otherwise), whereas the reference of ‘is a Republican’ is a function which returns the truth-value ‘true’ when given as input an object which is a member of the Republican party (and the truth-value ‘false’ otherwise). This is what explains the fact that (3) is true and (4) false: Obama is a member of the Democratic party, and is not a member of the Republican party.

Matters get more complicated, and more controversial, as we extend this sort of theory of reference to cover more and more of the types of expressions we find in natural languages like English. (For an introduction, see Heim and Kratzer (1998).) But the above is enough to give a rough idea of how one might proceed. For example, some predicates, like ‘loves’ combine with two names to form a sentence, rather than one. So the reference of two-place predicates of this sort must be something which combines with a pair of objects to determine a truth-value — perhaps, that function from ordered pairs of objects to truth-values which returns the truth-value ‘true’ when given as input a pair of objects whose first member loves the second member, and ‘false’ otherwise.

2.1.2 *Theories of reference vs. semantic theories*

So let’s suppose that we have a theory of reference for a language, in the above sense. Would we then have a satisfactory semantic theory for the language?

Some plausible arguments indicate that we would not. To adopt an example from Quine (1970), consider the pair of sentences

(5) All cordates are cordates.

(6) All cordates are renates.

Let’s suppose that both are true. If they are, then it looks as though that all cordates (creatures with a heart) are also renates (creatures with a kidney). Then it looks, from the point of view of the theory of reference, that (5) and (6) are just the same: they differ only in the substitution of ‘renates’ for ‘cordates’, and these expressions have the same reference (because they stand for the same function from objects to truth-values).

All the same, there is clearly an intuitive difference in meaning between (5) and (6); the sentences seem, in some sense, to say different things. The first seems to express the trivial, boring thought that every creature with a heart is a creature with a heart, whereas the second expresses the non-trivial, potentially informative claim that every creature with a heart also has a kidney. This suggests that there is an important difference between (5) and (6) which our theory of reference simply fails to capture.

Examples of the same sort can be generated using pairs of expressions of other types which share a reference, but intuitively differ in meaning; for example, ‘Clark Kent’ and ‘Superman,’ or (an example famously discussed by Frege (1892/1960)), ‘the Morning Star’ and ‘the Evening Star.’

This might seem a rather weak argument for the incompleteness of the theory of reference, resting as it does on intuitions about the relative informativeness of sentences like (5) and (6). But this argument can be strengthened by embedding sentences like (5) and (6) in more complex sentences, as follows:

(7) John believes that all cordates are cordates.

(8) John believes that all cordates are renates.

(7) and (8) differ only with respect to the underlined expressions and, as we noted above, these expressions have the same reference. Despite this, it seems clear that (7) and (8) could differ in truth-value: someone could know that all cordates have a heart without having any opinion on the question of whether all renates do. But that means that the references of expressions don’t even do the job for which they were introduced: they don’t explain the contribution that expressions make to the determination of the truth-value of all sentences in which they occur. If we are to be able to explain, in terms of the properties of the expressions that make them up, how (7) and (8) can differ in truth-value, then expressions must have some other sort of value, some sort of meaning, which goes beyond reference.

(7) and (8) are called *belief ascriptions*, for the obvious reason that they ascribe a belief to a subject. Belief ascriptions are one sort of *propositional attitude ascription* — other types include ascriptions of knowledge, desire, or judgement. As will become clear in what follows, propositional attitude ascriptions have been very important in recent debates in semantics. One of the reasons why they have been important is exemplified by (7) and (8). Because these sentences can differ in truth-value despite the fact that they differ only with respect to the underlined words, and these words both share a reference and occupy the same place in the structure of the two sentences, we say that (7) and (8) contain a *non-extensional context*: roughly, a ‘location’ in the sentence which is such that substitution of terms which share a reference in that location can change truth-value. (They’re called ‘non-extensional contexts’ because ‘extension’ is another term for ‘reference.’)

We can give a similar argument for the incompleteness of the theory of reference based on the substitution of whole sentences. A theory of reference assigns to subsentential expressions values which explain their contribution to the truth-values of sentences; but to those sentences, it only assigns ‘true’ or ‘false.’ But consider a pair of sentences like

(9) Mary believes that Barack Obama is the president of the United States.

(10) Mary believes that John Key is the prime minister of New Zealand.

Because both of the underlined sentences are true, (9) and (10) are a pair of sentences which differ only with respect to substitution of expressions (namely, those sentences) with the same reference. Nonetheless, (9) and (10) could plainly differ in truth-value.

This seems to show that a semantic theory should assign some value to sentences other than a truth-value. Another route to this conclusion is the apparent truth of claims of the following sort:

There are three things that John believes about Indiana, and they are all false.

There are many necessary truths which are not a priori, and my favorite sentence expresses one of them.

To get an A you must believe everything I say.

Sentences like these seem to show that there are things which are the objects of mental states like belief, the bearers of truth and falsity, and the things expressed by sentences. What are these things? The theory of reference provides no answer.

Friends of propositions aim both to provide a theory of these entities, and, in so doing, also to solve the two problems for the theory of reference discussed above: (i) the lack of an explanation for the fact that (5) is trivial and a priori while (6) is not, and (ii) the fact (exemplified by (7)/(8) and (9)/(10)) that sentences which differ only in the substitution of expressions with the same reference can differ in truth-value.

A theory of propositions thus does not abandon the theory of reference, as sketched above, but simply says that there is more to a semantic theory than the theory of reference. Subsentential expressions have, in addition to a reference, a *content*. The contents of sentences — what sentences express — are known as *propositions*.

2.1.3 *The relationship between content and reference*

The natural next question asks what sorts of things these contents are; below I'll discuss some of the leading views about this. But in advance of laying out any theory about what contents are, we can say some general things about the role that contents are meant to play.

First, what is the relationship between content and reference? Let's examine this question in connection with sentences; here it amounts to the question of the relationship between the proposition a sentence expresses and the sentence's truth-value. One point brought out by the example of (9) and (10) is that two sentences can express different propositions while having the same truth-value. After all, the beliefs ascribed to Mary by these sentences are different; so if

propositions are the objects of belief, the propositions corresponding to the underlined sentences must be different. Nonetheless, both sentences are true.

Is the reverse possible? Can two sentences express the same proposition, but differ in truth-value? It seems not, as can be illustrated again by the role of propositions as the objects of belief. Suppose that you and I believe the exact same thing — both of us believe the world to be just the same way. Can my belief be true, and yours false? It seems not; it is incoherent to say that we both believe the world to be the same way, but that I get things right and you get them wrong. So it seems that if two sentences express the same proposition, they must have the same truth value.

In general, then, it seems plausible that two sentences with the same content — i.e., which express the same proposition — must always have the same reference, though two expressions with the same reference can differ in content. This is the view stated by the Fregean slogan that sense determines reference ('sense' being the conventional translation of Frege's *Sinn*, which was his word for what we are calling 'content').

If this holds for sentences, does it also hold for subsentential expressions? It seems that it must. Suppose for reductio that two subsentential expressions, e_1 and e_2 , have the same content but differ in reference. It seems plausible that two sentences which differ only by the substitution of expressions with the same content must have the same content. (While plausible, this principle is not uncontroversial; see COMPOSITIONALITY.) But if this is true, then sentences which differ only in the substitution of e_1 and e_2 would have the same content. But such a pair of sentences could differ in truth-value, since, for any pair of expressions which differ in reference, there is some pair of sentences which differ only by the substitution of those expressions and differ in truth-value. So if there could be a pair of expressions like e_1 and e_2 , which differ in content but not reference, there could be a pair of sentences which have the same content — which express the same proposition — but differ in truth-value. But this is impossible; hence there could be no pair of expressions like e_1 and e_2 , and content must determine reference for subsentential expressions as well as sentences.

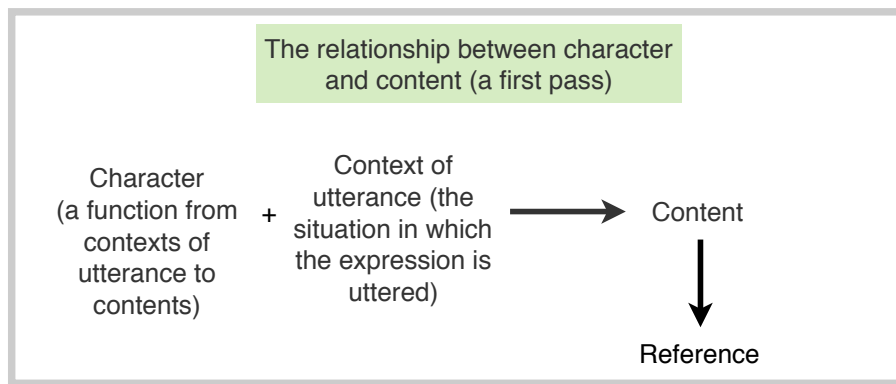
This result — that content determines reference — explains one thing we should, plausibly, want a semantic theory to do: it should assign to each expression some value — a content — which determines a reference for that expression.

2.1.4 *Character and content, context and circumstance*

However, there is an obvious problem with the idea that we can assign a content, in this sense, to all of the expressions of a language like English: many expressions, like 'I' or 'here', have a different reference when uttered by different speakers in different situations. So we plainly cannot assign to 'I' a single content which determines a reference for the expression, since the expression has a different reference in different situations. These 'situations' are typically called *contexts of utterance*, or just *contexts*, and expressions whose reference depends on the context are called INDEXICALS or *context-dependent expressions*.

The obvious existence of such expressions shows that a semantic theory must

do more than simply assign contents to every expression of the language. Expressions like ‘I’ must also be associated with rules which determine the content of the expression, given a context of utterance. These rules, which are functions from contexts to contents, are called *characters*. (The terminology here, as well as the view of the relationship between context, content, and reference, is due to Kaplan (1989).) So the character of ‘I’ must be some function from contexts to contents which delivers a content which determines me as reference in a context in which I am the speaker, delivers a content which determines Barack Obama as reference in a context in which Barack Obama is the speaker, and so on.



Here we face another potentially misleading ambiguity in ‘meaning.’ What is the real *meaning* of an expression — its character, or its content (in the relevant context)? This is an empty terminological question. Expressions have characters which, given a context, determine a content. We can talk about either character or content, and both are important. Nothing is to be gained by arguing that one rather than the other deserves the title of ‘meaning.’ The important thing is to be clear on the distinction, and to see the reasons for thinking that expressions have both a character and (relative to a context) a content.

How many indexical expressions are there? There are some obvious candidates — ‘I’, ‘here’, ‘now’, etc. — but beyond the obvious candidates, it is very much a matter of dispute. (For an opinionated overview, see Cappelen and LePore (2005).) But there is a kind of argument which seems to show that almost every expression is an indexical. Consider an expression which does not seem to be context-sensitive, like ‘the second-largest city in the United States.’ This does not seem to be context-sensitive, because it seems to refer to the same city — Los Angeles — whether uttered by me, you, or some other English speaker. But now consider a sentence like

- (11) 100 years ago, the second-largest city in the United States was Chicago.

This sentence is true. But for it to be true, ‘the second-largest city in the United States’ would have to, in (11), refer to Chicago. But then it seems like this expression must be an indexical — its reference must depend on the context of utterance. In (11), the thought goes, the phrase ‘one hundred years ago’ shifts the context: in (11), ‘the second-largest city in the United States’ refers to that city that it would have referred to if uttered one hundred years ago.

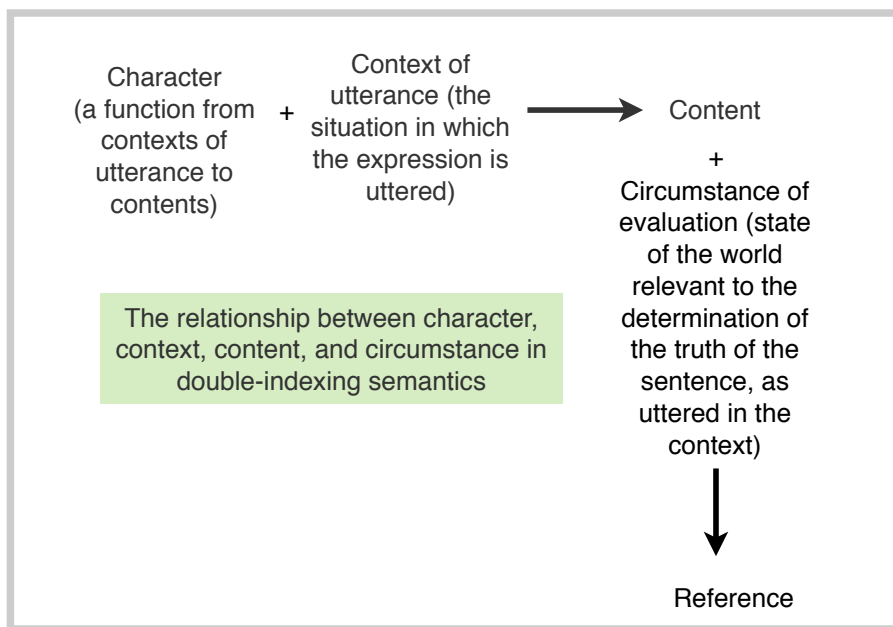
However, this can’t be quite right, as is shown by examples like this one:

(12) In 100 years, I will not exist.

Let’s suppose that this sentence, as uttered by me, is true. Then, if what we said about (11) was right, it seems that ‘I’ must, in, (12), refer to whoever it would refer to if it were uttered 100 years in the future. So the one thing we know is that (assuming that (12) is true), it does not refer to me — after all, I won’t be around to utter anything. But, plainly, the ‘I’ in (12) *does* refer to me when this sentence is uttered by me — after all, it is a claim about me. What’s going on here?

What examples like (12) show is that the reference of an expression must be relativized, not just to a context of utterance, but also to a *circumstance of evaluation* — roughly, the possible state of the world relevant to the determination of the truth or falsity of the sentence. In the case of many simple sentences, context and circumstance coincide; details aside, they both just are the state of the world at the time of the utterance, with a designated speaker and place. But sentences like (12) show that they can come apart. Phrases like ‘In 100 years’ shift the circumstance of evaluation — they change the state of the world relevant to the evaluation of the truth or falsity of the sentence — but don’t change the context of utterance. That’s why when I utter (12), ‘I’ refers to me — despite the fact that I won’t exist to utter it in 100 years time.

This is sometimes called the need for *double-indexing semantics* — the two indices being contexts of utterance and circumstances of evaluation.



The classic explanation of a double-indexing semantics is Kaplan (1989); another important early discussion is Kamp (1971). For a different interpretation of the framework, see Lewis (1980).

Double-indexing explains how we can regard the reference of ‘the second-largest city in the United States’ in (11) to be Chicago, without taking ‘the second-largest city in the United States’ to be an indexical like ‘I.’ On this view, ‘the second-largest city in the United States’ does not vary in content depending on the context of utterance; rather, the content of this phrase is such that it determines a different reference with respect to different circumstances of evaluation. In particular, it has Los Angeles as its reference with respect to the present state of the actual world, and has Chicago as its reference with respect to the state of actual world 100 years ago, in 1910.² Because ‘the second-largest city in the United States’ refers to different things with respect to different circumstances, it is not a RIGID DESIGNATOR — these being expressions which refer to the same object with respect to every circumstance of evaluation at which that object exists, and never refer to anything else with respect to another circumstance of evaluation. (The term ‘rigid designator’ is due to Kripke (1972).)

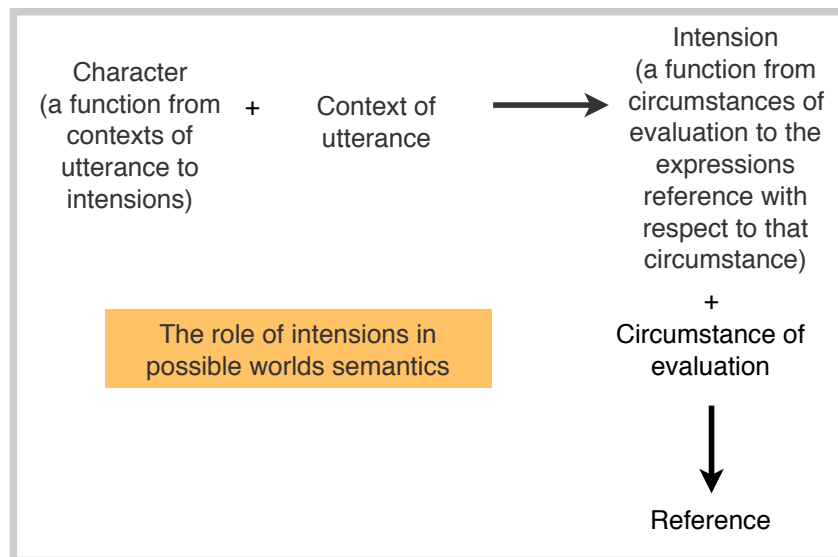
²This way of putting the point assumes that contents can have different references with respect to different times; this is a controversial assumption discussion of which would take us too far afield. (For an argument against this assumption, see Richard (1981).) Even if this assumption is false, an argument for double-indexing semantics parallel to the above could be constructed using changes in reference across worlds rather than across times.

2.1.5 Possible worlds semantics

So we know that expressions are associated with characters, which are functions from contexts to contents; and we know that contents are things which, for each circumstance of evaluation, determine a reference. We can now raise the central question of (propositional) semantic theories: what sorts of things are contents? The foregoing suggests a pleasingly minimalist answer to this question: perhaps, since contents are things which together with circumstances of evaluation determine a reference, contents just *are* functions from circumstances of evaluation to a reference.

This view sounds abstract but is, in a way, quite intuitive. The idea is that the meaning of an expression is not what the expression stands for in the relevant circumstance, but rather a rule which tells you what the expression would stand for were the world a certain way. So, on this view, the content of an expression like ‘the tallest man in the world’ is not simply the man who happens to be tallest, but rather a function from ways the world might be to men — namely, that function which, for any way the world might be, returns as a referent the tallest man in that world (if there is one, and nothing otherwise). This fits nicely with the intuitive idea that to understand such an expression one needn’t know what the expression actually refers to — after all, one can understand ‘the tallest man’ without knowing who the tallest man is — but must know how to tell what the expression would refer to, given certain information about the world (namely, the heights of all the men in it).

These functions, or rules, are called (following Carnap (1947)) *intensions*. Possible worlds semantics is the view that contents are intensions (and hence that characters are functions from contexts to intensions, i.e. functions from contexts to functions from circumstances of evaluation to a reference). (‘Intension’ is sometimes used more generally, as a synonym for ‘content.’ This usage is misleading, and the term is better reserved for functions from contexts to referents. It is then controversial whether, as the proponent of possible worlds semantics thinks, contents are intensions.)



For discussion of the application of the framework of possible world semantics to natural language, see Lewis (1970).

The intension of a sentence — i.e., the proposition that sentence expresses, on the present view — will then be a function from worlds to truth-values. In particular, it will be that function which returns the truth-value ‘true’ for every world with respect to which that sentence is true, and ‘false’ otherwise. The intension of a simple predicate like ‘is red’ will be a function from worlds to the function from objects to truth-values which, for each world, determines the truth-value ‘true’ if the thing in question is red, and false otherwise. In effect, possible worlds semantics takes the meanings of expressions to be functions from worlds to the values which would be assigned by a theory of reference to those expressions at the relevant world: in that sense, intensions are a kind of ‘extra layer’ on top of the theory of reference.

This extra layer promises to solve the problem posed by non-extensional contexts, as illustrated by the example of ‘cordate’ and ‘renate’ in (7) and (8). Our worry was that, since these expressions have the same reference, if meaning just is reference, then it seems that any pair of sentences which differ only in the substitution of these expressions must have the same truth-value. But (7) and (8) are such a pair of sentences, and needn’t have the same truth-value. The proponent of possible worlds semantics solves this problem by identifying the meaning of these expressions with their intension rather than their reference, and by pointing out that ‘cordate’ and ‘renate’, while they share a reference, seem to have different intensions. After all, even if in our world every creature with a heart is a creature with a kidney (and vice versa), it seems that the world could have been such that some creatures had a heart but not a kidney. Since with respect to that circumstance of evaluation the terms will differ in reference, their intensions — which are just functions from circumstances of evaluations

to referents — must also differ. Hence possible worlds semantics leaves room for (7) and (8) to differ in truth value, as they manifestly can.

The central problem facing possible worlds semantics, however, concerns sentences of the same form as (7) and (8): sentences which ascribe propositional attitudes, like beliefs, to subjects. To see this problem, we can begin by asking: according to possible worlds semantics, what does it take for a pair of sentences to have the same content (i.e., express the same proposition)? Since contents are intensions, and intensions are functions from circumstances of evaluation to referents, it seems that two sentences have the same content, according to possible worlds semantics, if they have the same truth-value with respect to every circumstance of evaluation. In other words, two sentences express the same proposition if and only if it is impossible for them to differ in truth-value.

The problem is that there are sentences which have the same truth-value in every circumstance of evaluation, but seem to differ in meaning. Consider, for example

(13) $2+2=4$.

(14) There are infinitely many prime numbers.

(13) and (14) are in a way reminiscent of (5) and (6): the first seems to be a triviality which everyone knows, and the second seems to be a more substantial claim of which one might well be ignorant. However, both are necessary truths: like any truths of mathematics, neither depends on special features of the actual world, but is rather true with respect to every circumstance of evaluation. Hence (13) and (14) have the same intension and, according to possible worlds semantics, must have the same content.

This is highly counterintuitive. The problem (just as with (5) and (6)) can be sharpened by embedding these sentences in propositional attitude ascriptions:

(15) John believes that $2+2=4$.

(16) John believes that there are infinitely many prime numbers.

As we have just seen, the proponent of possible worlds semantics must take the underlined sentences, (13) and (14), to have the same content; hence the proponent of possible worlds semantics must take (15) and (16) to be a pair of sentences which differ only in the substitution of expressions with the same content. But then the proponent of possible worlds semantics must take this pair of sentences to express the same proposition, and have the same truth-value; but (15) and (16) (like (7) and (8)) clearly can differ in truth-value, and hence clearly do not express the same proposition.

Indeed, the problem, as shown in Soames (1988), is worse than this. Consider a pair of sentences like

(17) Grass is green.

(18) Grass is green and there are infinitely many prime numbers.

The second of these is just the first conjoined with a necessary truth; hence the second is true if and only if the first is true. But then they have the same intension and, according to possible worlds semantics, have the same content. Hence the following two sentences cannot differ in truth-value:

(19) John believes that grass is green.

(20) John believes that grass is green and there are infinitely many prime numbers.

since they differ only by the substitution of expressions with the same content. Furthermore, it seems that belief distributes over conjunction, in this sense: anyone who believes the conjunction of a pair of propositions must also believe each of those propositions. But then if (20) is true, so must be (19). So it follows that (19) implies (20), and anyone who believes that grass is green must also believe that there are infinitely many primes. This line of argument generalizes to show that anyone who believes any propositions at all must believe every necessary truth. This is, at best, a highly counterintuitive consequence of the possible worlds semanticist's equation of contents with intensions. All things being equal, it seems that we should seek an approach to semantics which does not have this consequence. (For an attempt to reply to the argument from within the framework of possible worlds semantics, see Stalnaker (1984); for discussion of a related approach to semantics which aims to avoid these problems, see SITUATIONS IN NATURAL LANGUAGE SEMANTICS.)

2.1.6 Russellian propositions

What we need, then, is an approach to semantics which can explain how sentences like (13) and (14), and hence also (15) and (16), can express different propositions. That is, we need a view of propositions which makes room for the possibility that a pair of sentences can be true in just the same circumstances but nonetheless have genuinely different contents.

A natural thought is that (13) and (14) have different contents because they are about different things; for example, (14) makes a general claim about the set of prime numbers whereas (13) is about the relationship between the numbers 2 and 4. One might want our semantic theory to be sensitive to such differences: to count two sentences as expressing different propositions if they have different subject matters, in this sense. One way to secure this result is to think

of the contents of subsentential expressions as components of the proposition expressed by the sentence as a whole. Differences in the contents of subsentential expressions would then be sufficient for differences in the content of the sentence as a whole; so, for example, since (14) but not (13) contains an expression which refers to barbers, these sentences will express different propositions.

Proponents of this sort of view think of propositions as *structured*: as having constituents which include the meanings of the expressions which make up the sentence expressing the relevant proposition. (See, for more discussion, STRUCTURED PROPOSITIONS.) One important question for views of this sort is: what does it mean for an abstract object, like a proposition, to be structured, and have constituents? But this question would take us too far afield into metaphysics. The fundamental *semantic* question for proponents of this sort of structured proposition view is: what sorts of things are the constituents of propositions?

The answer to this question given by a proponent of *Russellian propositions* is: objects, properties, relations, and functions. (The view is called ‘Russellianism’ because of its resemblance to the view of content defended in Chapter IV of Russell (1903).) So described, Russellianism is a general view about what sorts of things the constituents of propositions are, and does not carry a commitment to any views about the contents of particular types of expressions. However, most Russellians also endorse a particular view about the contents of proper names which is known as *Millianism*: the view that the meaning of a simple proper name is the object (if any) for which it stands.

Russellianism has much to be said for it. It not only solves the problems with possible worlds semantics discussed above, but fits well with the intuitive idea that the function of names is to single out objects, and the function of predicates is to (what else?) predicate properties of those objects.

However, Millian-Russellian semantic theories also face some problems. Some of these are metaphysical in nature, and are based on the fact that propositions which have objects among their constituents cannot exist in circumstances in which those objects do not exist. (For discussion, see SINGULAR PROPOSITIONS, §§4-5.) Of the semantic objections to Millian-Russellian semantics, two are especially important.

The first of these problems involves the existence of *empty names*: names which have no referent. It is a commonplace that there are such names; an example is ‘Vulcan,’ the name introduced for the planet between Mercury and the sun which was causing perturbations in the orbit of Mercury. Because the Millian-Russellian says that the content of a name is its referent, the Millian-Russellian seems forced into saying that empty names lack a content. But this is surprising; it seems that we can use empty names in sentences to express propositions and form beliefs about the world. The Millian-Russellian owes some explanation of how this is possible, if such names genuinely lack a content. (For discussion, see Braun (1993).)

Perhaps the most important problem facing Millian-Russellian views, though, is Frege’s puzzle. Consider the sentences

(21) Clark Kent is Clark Kent.

(22) Clark Kent is Superman.

According to the Millian-Russellian, (21) and (22) differ only in the substitution of expressions which have the same content: after all, ‘Clark Kent’ and ‘Superman’ are proper names which refer to the same object, and the Millian-Russellian holds that the content of a proper name is the object to which that name refers. But this is a surprising result. These sentences seem to differ in meaning, because (21) seems to express a trivial, obvious claim, whereas (22) seems to express a non-trivial, potentially informative claim.

This sort of objection to Millian-Russellian views can (as above) be strengthened by embedding the intuitively different sentences in propositional attitude ascriptions, as follows:

(23) Lois believes that Clark Kent is Clark Kent.

(24) Lois believes that Clark Kent is Superman.

The problem posed by (23) and (24) for Russellian semantics is analogous to the problem posed by (15) and (16) for possible worlds semantics. Here, as there, we have a pair of belief ascriptions which seem as though they could differ in truth-value despite the fact that these sentences differ only with respect to expressions counted as synonymous by the relevant semantic theory.

Russellians have offered a variety of responses to Frege’s puzzle. Many Russellians think that our intuition that sentences like (23) and (24) can differ in truth-value is based on a mistake. This mistake might be explained at least partly in terms of a confusion between the proposition semantically expressed by a sentence in a context and the propositions speakers would typically use that sentence to pragmatically convey (Salmon (1986) and Soames (2002)), or in terms of the fact that a single proposition may be believed under several ‘propositional guises’ (again, see Salmon (1986)), or in terms of a failure to integrate pieces of information stored using distinct mental representations (Braun and Saul (2002)).³ Alternatively, a Russellian might try to make room for (23) and (24) to genuinely differ in truth-value by giving up the idea that sentences which differ only in the substitution of proper names with the same content must express the same proposition (Fine (2007)).

³According to the first sort of Russellian, we can’t adequately decide disputes about the semantics of names without also investigating the relationship between semantics and PRAGMATICS; one of the important trends in recent semantics has been an increasing attempt to explain linguistic phenomena with an eye on both possible semantic and pragmatic explanations.

2.1.7 Fregean propositions

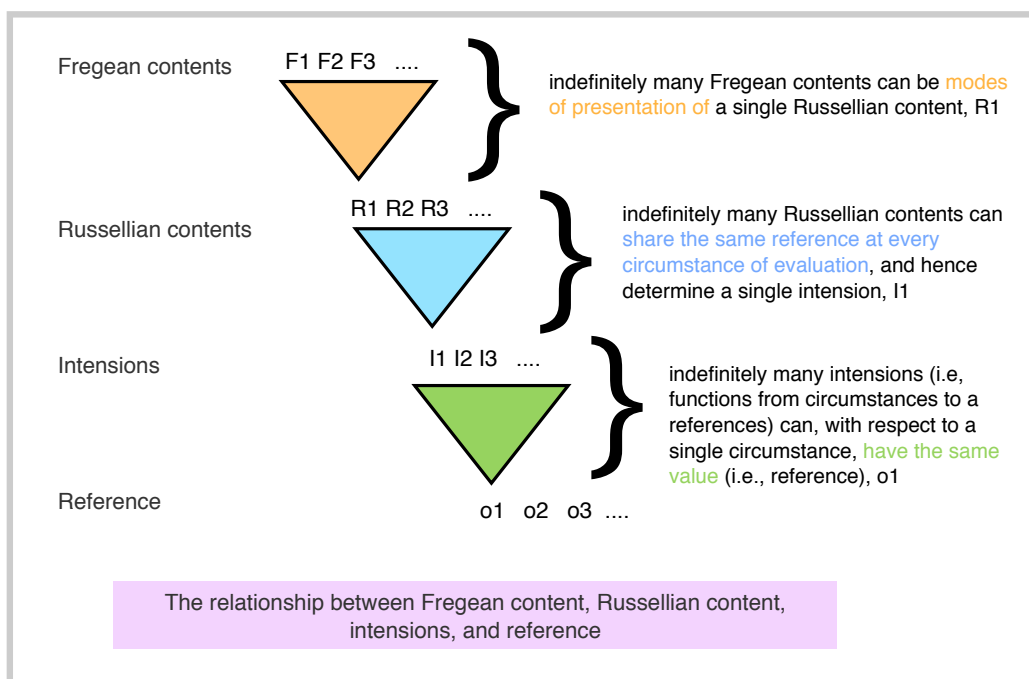
However, these are not the only responses to Frege's puzzle. Just as the Russellian responded to the problem posed by (15) and (16) by holding that two sentences with the same intension can differ in meaning, one might respond to the problem posed by (23) and (24) by holding that two names which refer to the same object can differ in meaning, thus making room for (23) and (24) to differ in truth-value. This is to endorse a Fregean response to Frege's puzzle, and to abandon the Russellian approach to semantics (or, at least, to abandon Millian-Russellian semantics).

Fregeans, like Russellians, think of the proposition expressed by a sentence as a structured entity with constituents which are the contents of the expressions making up the sentence. But Fregeans, unlike Russellians, do not think of these propositional constituents as the objects, properties, and relations for which these expressions stand; instead, Fregeans think of the contents as modes of presentation, or ways of thinking about, objects, properties, and relations. The standard term for these modes of presentation is *sense*. (As with 'intension,' 'sense' is sometimes also used as a synonym for 'content.' But, as with 'intension,' it avoids confusion to restrict 'sense' for 'content, as construed by Fregean semantics.' It is then controversial whether there are such things as senses, and whether they are the contents of expressions.) Frege explained his view of senses with an analogy:

“The reference of a proper name is the object itself which we designate by its means; the idea, which we have in that case, is wholly subjective; in between lies the sense, which is indeed no longer subjective like the idea, but is yet not the object itself. The following analogy will perhaps clarify these relationships. Somebody observes the Moon through a telescope. I compare the Moon itself to the reference; it is the object of the observation, mediated by the real image projected by the object glass in the interior of the telescope, and by the retinal image of the observer. The former I compare to the sense, the latter is like the idea or experience. The optical image in the telescope is indeed one-sided and dependent upon the standpoint of observation; but it is still objective, inasmuch as it can be used by several observers. At any rate it could be arranged for several to use it simultaneously. But each one would have his own retinal image.” (Frege (1892/1960))

Senses are then objective, in that more than one person can express thoughts with a given sense, and correspond many-one to objects. Thus, just as Russellian propositions correspond many-one to intensions, Fregean propositions correspond many-one to Russellian propositions. This is sometimes expressed by the claim that Fregean contents are more *fine-grained* than Russellian contents (or intensions).

Indeed, we can think of our three propositional semantic theories, along with the theory of reference, as related by this kind of many-one relation:



The principle argument for Fregean semantics (which also motivated Frege himself) is the neat solution the view offers to Frege’s puzzle: the view says that, in cases like (23) and (24) in which there seems to be a difference in content, there really is a difference in content: the names share a reference, but differ in their sense, because they differ in their mode of presentation of their shared reference.

The principal challenge for Fregeanism is the challenge of giving a non-metaphorical explanation of the nature of sense. This is a problem for the Fregean in a way that it is not for the possible worlds semanticist or the Russellian since the Fregean, unlike these two, introduces a new class of entities to serve as meanings of expressions rather than merely appropriating an already recognized sort of entity — like a function, or an object, property, or relation — to serve this purpose.⁴ A first step toward answering this challenge is provided by a criterion for telling when two expressions differ in meaning, which might be stated as follows. In his 1906 paper, ‘A Brief Survey of My Logical Doctrines,’ Frege seems to endorse the following criterion:

Frege’s criterion of difference for senses

Two sentences S and S' differ in sense if and only if some rational

⁴This is not to say that there is not controversy over the existence of these sorts of entities; there is. (See for discussion NOMINALISM IN METAPHYSICS.) The point is just that many believe in the existence of entities of these kinds for reasons unconnected to the philosophy of language; these entities were not introduced to play the role of the contents of expressions.

agent who understood both could, on reflection, judge that S is true without judging that S' is true.

One worry about this formulation concerns the apparent existence of pairs of sentences, like ‘If Obama exists, then Obama=Obama’ and ‘If McCain exists, McCain=McCain’ which are such that any rational person who understands both will take both to be true. These sentences seem intuitively to differ in content — but this is ruled out by the criterion above. One idea for getting around this problem would be to state our criterion of difference for senses of expressions in terms of differences which result from substituting one expression for another:

Two expressions e and e' differ in sense if there are a pair of sentences, S and S' which (i) differ only in the substitution of e for e' and (ii) are such that some rational agent who understood both could, on reflection, judge that S is true without judging that S' is true.

This version of the criterion has Frege’s formulation as a special case, since sentences are, of course, expressions; and it solves the problem with obvious truths, since it seems that substitution of sentences of this sort can change the truth value of a propositional attitude ascription. Furthermore, the criterion delivers the wanted result that coreferential names like ‘Superman’ and ‘Clark Kent’ differ in sense, since a rational, reflective agent like Lois Lane could think that (21) is true while withholding assent from (22).

But even if this tells us when names differ in sense, it does not quite tell us what the sense of a name *is*. Here is one initially plausible way of explaining what the sense of a name is. We know that, whatever the content of a name is, it must be something which determines as a reference the object for which the name stands; and we know that, if Fregeanism is true, this must be something other than the object itself. A natural thought, then, is that the content of a name — its sense — is some condition which the referent of the name uniquely satisfies. Coreferential names can differ in sense because there is always more than one condition which a given object uniquely satisfies. (For example, Superman/Clark Kent uniquely satisfies both the condition of being the superhero Lois most admires, and the newspaperman she least admires.) Given this view, it is natural to then hold that names have the same meanings as *definite descriptions* — phrases of the form ‘the so-and-so.’ After all, phrases of this sort seem to be designed to pick out the unique object, if any, which satisfies the condition following the ‘the.’ (For more discussion, see DESCRIPTIONS.) This Fregean view of names is called *Fregean descriptivism*.

However, as Saul Kripke argued in *Naming and Necessity*, Fregean descriptivism faces some serious problems. Here is one of the arguments he gave against the view, which is called the *modal argument*. Consider a name like ‘Aristotle,’ and suppose for purposes of exposition that the sense I associate with that name is the sense of the definite description ‘the greatest philosopher of antiquity.’ Now consider the following pair of sentences:

(25) Necessarily, if Aristotle exists, then Aristotle is Aristotle.

(26) Necessarily, if Aristotle exists, then Aristotle is the greatest philosopher of antiquity.

If Fregean descriptivism is true, and ‘the greatest philosopher of antiquity’ is indeed the description I associate with the name ‘Aristotle,’ then it seems that (25) and (26) must be a pair of sentences which differ only via the substitution of expressions (the underlined ones) with the same content. If this is right, then (25) and (26) must express the same proposition, and have the same truth-value. But this seems to be a mistake; while (25) appears to be true (Aristotle could hardly have failed to be himself), (26) appears to be false (perhaps Aristotle could have been a shoemaker rather than a philosopher; or perhaps if Plato had worked a bit harder, he rather than Aristotle could have been the greatest philosopher of antiquity).

Fregean descriptivists have given various replies to Kripke’s modal and other arguments; see especially Plantinga (1978); Dummett (1981); Sosa (2001). For rejoinders to these Fregean replies, see Soames (1998, 2002); Caplan (2005). For a brief sketch of Kripke’s other arguments against Fregean descriptivism, see NAMES, §2.4.

Kripke’s arguments provide a strong reason for Fregeans to deny Fregean descriptivism, and hold instead that the senses of proper names are not the senses of any definite description associated with those names by speakers. The main problem for this sort of *non-descriptive Fregeanism* is to explain what the sense of a name might be such that it can determine the reference of the name, if it is not a condition uniquely satisfied by the reference of the name. Non-descriptive Fregean views are defended in McDowell (1977) and Evans (1981); for a version of the view which gives up the idea that the sense of a name determines its reference, see Chalmers (2004, 2006).

Two other problems for Fregean semantics are worth mentioning. The first calls into question the Fregean’s claim to have provided a plausible solution to Frege’s puzzle. The Fregean resolves instances of Frege’s puzzle by positing differences in sense to explain apparent differences in truth-value. But this sort of solution, if pursued generally, seems to lead to the surprising result that no two expressions can have the same content. For consider a pair of expressions which really do seem to have the same content, like ‘catsup’ and ‘ketchup.’ (The example, as well as the argument to follow, are borrowed from Salmon (1990).) Now consider Bob, a rather confused condiment user, who thinks that the tasty red substance standardly labeled ‘catsup’ is distinct from the tasty red substance standardly labeled ‘ketchup’, and consider the following pair of sentences:

(27) Bob believes that catsup is catsup.

(28) Bob believes that catsup is ketchup.

(27) and (28) seem quite a bit like (23) and (24): these each seem to be pairs of sentences which differ in truth-value, despite differing only in the substitution of the underlined expressions. So, for consistency, it seems that the Fregean should explain the apparent difference in truth-value between (27) and (28) in just the way he explains the apparent difference in truth-value between (23) and (24): by positing a difference in meaning between the underlined expressions. But, first, it is hard to see how expressions like ‘catsup’ and ‘ketchup’ could differ in meaning; and, second, it seems that an example of this sort could be generated for any alleged pair of synonymous expressions. (A closely related series of examples is developed in much more detail in Kripke (1979).)

The example of ‘catsup’ and ketchup’ is related to a second worry for the Fregean, which is the reverse of the Fregean’s complaint about Russellian semantics: a plausible case can be made that Frege’s criterion of difference for sense slices contents too finely, and draws distinctions in content where there are none. One way of developing this sort of argument involves (again) propositional attitude ascriptions. It seems plausible that if I utter a sentence like ‘Hammurabi thought that Hesperus was visible only in the morning,’ what I say is true if and only if one of Hammurabi’s thoughts has the same content as does the sentence ‘Hesperus was visible only in the morning,’ as used by me. On a Russellian view, this places a reasonable constraint on the truth of the ascription; it requires only that Hammurabi believe of a certain object that it instantiates the property of being visible in the morning. But on a Fregean view, this sort of view of attitude ascriptions would require that Hammurabi thought of the planet Venus under the same mode of presentation as I attach to the term ‘Hesperus.’ This seems implausible, since it seems that I can truly report Hammurabi’s beliefs without knowing anything about the mode of presentation under which he thought of the planets. (For a recent attempt to develop a Fregean semantics for propositional attitude ascriptions which avoids this sort of problem, see Chalmers (2007).)

2.2 *Non-propositional theories*

So, while there are powerful motivations for propositional semantic theories, each theory of this sort also faces some difficult challenges. These challenges have led some to think that the idea behind propositional semantics — the idea that the job of a semantic theory is to systematically pair expressions with the entities which are their meanings — is fundamentally misguided. Wittgenstein was parodying just this idea when he wrote “You say: the point isn’t the word, but its meaning, and you think of the meaning as a thing of the same kind as the word, though also different from the word. Here the word, there the meaning. The money, and the cow that you can buy with it” (§120).

While Wittgenstein himself did not think that systematic theorizing about semantics was possible, this anti-theoretical stance has not been shared by all

subsequent philosophers who share his aversion to “meanings as entities.” This section is intended to provide some idea of how semantics might work in a framework which eschews propositions and their constituents, by explaining the basics of two representative theories within this tradition.

The difference between these theories is best explained by recalling the sort of theory of reference sketched in §2.1.1 above. Recall that propositional theories supplement this theory of reference with an extra layer — with a theory which assigns a content, as well as a reference, to each meaningful expression. One alternative to propositional theories — Davidsonian truth-conditional theories — takes this extra layer to be unnecessary, and holds that a theory of reference is all the semantic theory we need. A second, more radical alternative to propositional theories — Chomskyan internalist theories — holds not that a theory of reference is not enough, but rather that it is too much; on this view, the meanings of expressions of a natural language neither are, nor determine, a reference.

2.2.1 *The Davidsonian program*

One of the most important sources of opposition to the idea of “meanings as entities” is Donald Davidson. Davidson thought that semantic theory should take the form of a theory of truth for the language of the sort which Alfred Tarski showed us how to construct. (See Tarski (1936) and TARSKI’S TRUTH DEFINITIONS.)

For our purposes, it will be convenient to think of a Tarskian truth theory as a variant on the sorts of theories of reference introduced in §2.1.1. Recall that theories of reference of this sort specified, for each proper name in the language, the object to which that name refers, and for every simple predicate in the language, the set of things which satisfy that predicate. If we then consider a sentence which combines a proper name with such a predicate, like

Amelia sings.

the theory tells us what it would take for that sentence to be true: it tells us that this sentence is true if and only if the object to which ‘Amelia’ refers is a member of the set of things which satisfy the predicate ‘sings’ — i.e., the set of things which sing. So we can think of a full theory of reference for the language as implying, for each sentence of this sort, a *T-sentence* of the form

“Amelia sings” is *T* (in the language) if and only if Amelia sings.

Suppose now that we expand our theory of reference so that it implies a *T-sentence* of this sort for every sentence of the language, rather than just for simple sentences which result from combining a name and a monadic predicate. We would then have a Tarskian truth theory for our language. Tarski’s idea was that such a theory would define a truth predicate (*T*) for the language; Davidson, by contrast, thought that we find in Tarskian truth theories

“the sophisticated and powerful foundation of a competent theory of meaning” (Davidson (1967/1984)).

This claim is puzzling: why should a theory which issues T-sentences, but makes no explicit claims about meaning or content, count as a *semantic* theory? Davidson’s answer was that knowledge of such a theory would be sufficient to understand the language. If Davidson were right about this, then he would have a plausible argument that a semantic theory could take this form. After all, it is plausible that someone who understands a language knows the meanings of the expressions in the language; so, if knowledge of a Tarskian truth theory for the language were sufficient to understand the language, then knowledge of what that theory says would be sufficient to know all the facts about the meanings of expressions in the language, in which case it seems that the theory would state all the facts about the meanings of expressions in the language.

One advantage of this sort of approach to semantics is its parsimony: it makes no use of the intensions, Russellian propositions, or Fregean senses assigned to expressions by the propositional semantic theories discussed above. Of course, as we saw above, these entities were introduced to provide a satisfactory semantic treatment of various sorts of linguistic constructions, and one might well wonder whether it is possible to provide a Tarskian truth theory of the sort sketched above for a natural language without making use of intensions, Russellian propositions, or Fregean senses. The Davidsonian program obviously requires that we be able to do this, but it is still very much a matter of controversy whether a truth theory of this sort can be constructed. Discussion of this point is beyond the scope of this entry; one good way into this debate is through the debate about whether the Davidsonian program can provide an adequate treatment of propositional attitude ascriptions. See the discussion of the paratactic account and interpreted logical forms in PROPOSITIONAL ATTITUDE REPORTS. (For Davidson’s initial treatment of attitude ascriptions, see Davidson (1968); for further discussion see, among other places, Burge (1986); Schiffer (1987); LePore and Loewer (1989); Larson and Ludlow (1993); Soames (2002).)

Let’s set this aside, and assume that a Tarskian truth theory of the relevant sort can be constructed, and ask whether, given this supposition, this sort of theory would provide an adequate semantics. There are two fundamental reasons for thinking that it would not, both of which are ultimately due to Foster (1976). I will follow Larson and Segal (1995) by calling these the *extension problem* and the *information problem*.

The extension problem stems from the fact that it is not enough for a semantic theory whose theorems are T-sentences to yield true theorems; the T-sentence

“Snow is white” is *T* in English iff grass is green

is true, but tells us hardly anything about the meaning of “Snow is white.” Rather, we want a semantic theory to entail, for each sentence of the object language, exactly one *interpretive* T-sentence: a T-sentence such that the sentence used on its right-hand side gives the meaning of the sentence mentioned on its

left-hand side. Our theory must entail *at least one* such T-sentence for each sentence in the object language because the aim is to give the meaning of each sentence in the language; and it must entail *no more than one* because, if the theory had as theorems more than one T-sentence for a single sentence S of the object language, an agent who knew all the theorems of the theory would not yet understand S , since such an agent would not know which of the T-sentences which mention S was interpretive.

The problem is that it seems that any theory which implies at least one T-sentence for every sentence of the language will also imply more than one T-sentence for every sentence in the language. For any sentences p, q , if the theory entails a T-sentence

S is T in L iff p ,

then, since p is logically equivalent to $p \ \& \ \neg(q \ \& \ \neg q)$, the theory will also entail the T-sentence

S is T in L iff $p \ \& \ \neg(q \ \& \ \neg q)$,

which, if the first is interpretive, won't be. But then the theory will entail at least one non-interpretive T-sentence, and someone who knows the theory will not know which of the relevant sentences is interpretive and which not; such a person therefore would not understand the language.

The information problem is that, even if our semantic theory entails all and only interpretive T-sentences, it is not the case that knowledge of what is said by these theorems would suffice for understanding the object language. For, it seems, I can know what is said by a series of interpretive T-sentences without knowing that they are interpretive. I may, for example, know what is said by the interpretive T-sentence

“Londres est jolie” is T in French iff London is pretty

but still not know the meaning of the sentence mentioned on the left-hand side of the T-sentence. The truth of what is said by this sentence, after all, is compatible with the sentence used on the right-hand side being materially equivalent to, but different in meaning from, the sentence mentioned on the left. This seems to indicate that knowing what is said by a truth theory of the relevant kind is not, after all, sufficient for understanding a language. (For replies to these criticisms, see Davidson (1976), Larson and Segal (1995) and Kölbel (2001); for criticism of these replies, see Soames (1992) and Speaks (2006).)

2.2.2 Chomskyan internalist semantics

There is another alternative to propositional semantics which is at least as different from the Davidsonian program as that program is from various propositional views. This view is sometimes called ‘internalist semantics’ by contrast

with views which locate the semantic properties of expressions in their relation to elements of the external world. An internalist approach to semantics is associated with the work of Noam Chomsky (see especially Chomsky (2000)).

It is easy to say what this approach to semantics denies. The internalist denies an assumption common to all of the approaches above: the assumption that in giving the content of an expression, we are primarily specifying something about that expressions' relation to things in the world which that expression might be used to say things about. According to the internalist, expressions as such don't bear any semantically interesting relations to things in the world; names don't, for example, *refer* to the objects with which one might take them to be associated. Sentences are not true or false, and do not express propositions which are true or false; the idea that we can understand natural languages using a theory of reference as a guide is mistaken. On this sort of view, we occasionally use sentences to say true or false things about the world, and occasionally use names to refer to things; but this is just one aspect of our use of names and sentences, and is not a claim about the meanings of those expressions.

It is more difficult, in a short space, to say what the internalist says the meanings of linguistic expressions are. According to McGilvray (1998), “[t]he basic thesis is that meanings are contents intrinsic to expressions . . . and that they are defined and individuated by syntax, broadly conceived” (225). This description is sufficient to show the difference between this view of meaning and those sketched above: it is not just that the focus is not on the relationship between certain syntactic items and non-linguistic reality, but that, according to this view, syntactic and semantic properties of expressions are held to be inseparable. McGilvray adds that “[t]his unusual approach to meaning has few contemporary supporters,” which is probably true — though less so now than in 1998, when this was written. For defenses and developments of this view, see McGilvray (1998); Chomsky (2000); Pietroski (2003, 2005).

3 Foundational theories of meaning

We now turn to our second sort of ‘theory of meaning’: foundational theories of meaning, which are attempts to specify the facts in virtue of which expressions of natural languages come to have the semantic properties that they have.

The question which foundational theories of meaning try to answer is a common sort of question in philosophy. In the philosophy of ACTION we ask what the facts are in virtue of which a given piece of behavior is an intentional action; in questions about PERSONAL IDENTITY we ask what the facts are in virtue of which x and y are the same person; in ethics we ask what the facts are in virtue of which a given action is morally right or wrong. But, even if they are common enough, it is not obvious what the constraints are on answers to these sorts of questions, or when we should expect questions of this sort to have interesting answers.

Accordingly, one sort of approach to foundational theories of meaning is simply to deny that there is any true foundational theory of meaning. One

might be quite willing to endorse one of the semantic theories outlined above while also holding that facts about the meanings of expressions are *primitive*, in the sense that there is no systematic story to be told about the facts in virtue of which expressions have the meanings that they have. (See, for example, Johnston (1988).)

Most philosophers have not, however, taken this view, and have held that there must be some systematic account of the facts about language users in virtue of which their words have the semantic properties that they do. Typically, such philosophers aim to specify properties of expressions which are necessarily and sufficient for, and explanatorily prior to, their having a certain meaning.

In the last fifty years, the dominant view about the foundations of meaning in analytic philosophy has been the *mentalist* view that the meanings of expressions in public languages are to be explained in terms of the contents of the mental states of users of those languages. In what follows I'll first discuss these mentalist foundational theories of meaning before turning to alternatives to mentalism.

3.1 *Mentalist theories*

All mentalist theories of meaning have in common that they analyze one sort of representation — linguistic representation — in terms of another sort of representation — mental representation. For philosophers who are interested in explaining content, or representation, in non-representational terms, then, mentalist theories can only be a first step in the task of giving an ultimate explanation of the foundations of linguistic representation. The second, and more fundamental explanation would then come at the level of a theory of mental content. (For an overview of theories of this sort, see MENTAL REPRESENTATION and the essays in Stich and Warfield (1994).) Indeed, the popularity of mentalist theories of linguistic meaning, along with the conviction that content should be explicable in non-representational terms, is an important reason why so much attention has been focused on theories of mental representation over the last few decades.

Since mentalists aim to explain the nature of meaning in terms of the mental states of language users, mentalist theories may be divided according to which mental states they take to be relevant to the determination of meaning. The most well-worked out views on this topic are the Gricean view, which explains meaning in terms of the communicative intentions of language users, and the view that the meanings of expressions are fixed by conventions which pair sentences with certain beliefs. We will discuss these in turn, followed by a brief discussion of a third alternative to which a mentalist might turn.

3.1.1 *The Gricean program*

PAUL GRICE developed an analysis of meaning which can be thought of as the conjunction of two claims: (1) facts about what expressions mean are to be explained, or analyzed, in terms of facts about what speakers mean by utterances of them, and (2) facts about what speakers mean by their utterances can be

explained in terms of their intentions. These two theses comprise the ‘Gricean program’ for reducing meaning to the contents of the intentions of speakers.

To understand Grice’s view of meaning, it is important first to be clear on the distinction between the meaning, or content, of *linguistic expressions* — which is what semantic theories like those discussed in §2 aim to describe — and what *speakers* mean by utterances employing those expressions. This distinction can be illustrated by example. (See PRAGMATICS for more discussion.) Suppose that in response to a question about the restaurants in the city where I live, I say “Well, South Bend is not exactly New York City.” The meaning of this sentence is fairly clear: it expresses the (true) proposition that South Bend, Indiana is not identical to New York City. But what *I* mean by uttering this sentence is something more than this triviality: I mean by the utterance not just that South Bend is distinct from New York, but also that the quality of food on offer in South Bend is not comparable to that in New York. And this example utterance is in an important respect very typical: usually the propositions which speakers mean to convey by their utterances include the propositions expressed by the sentences they use, but also include other propositions. When we ask ‘What did you mean by that?’ we are usually not asking for the meaning of the sentence uttered.

The idea behind stage (1) of Grice’s theory of meaning is that of these two phenomena, speaker-meaning is the more fundamental: sentences and other expressions mean what they do because of what speakers mean by their utterances of those sentences. One powerful way to substantiate this claim would be to show that facts about speaker-meaning may be given an analysis which makes no use of facts about what expressions mean; and this is just what stage (2) of Grice’s analysis, to which we now turn, aims to provide.

Grice thought that speaker-meaning could be analyzed in terms of the communicative intentions of speakers — in particular, their intentions to cause beliefs in their audience. Though there are many different versions of this sort of Gricean analysis of speaker-meaning, the following is as good as any:⁵

- [G] a means p by uttering $x \equiv a$ intends in uttering x that
- (1) his audience come to believe p ,
 - (2) his audience recognize this intention, &
 - (3) (1) occur on the basis of (2)⁶

One way to see the intuitive motivation behind analyses like [G] is to begin with the idea that meaning something by an utterance is a matter of trying to convey one’s beliefs. Trying to convey one’s beliefs can be thought of as intending someone to share one’s beliefs; but, fairly clearly, you can intend by an action that someone form a belief that p without meaning p by your action. An example here might help. Suppose I turn to you and say, “You’re standing on my foot.” I intend that you hear the words I am saying; so I intend that you believe that I have said, “You’re standing on my foot.” But I do not mean by my utterance that I have said, “You’re standing on my foot.” That *is* my

⁵See Grice (1957, 1969).

utterance — what I mean by it is the proposition that you are standing on my foot, or that you should get off of my foot. I do not mean by my utterance that I am uttering a certain sentence.

This sort of example indicates that speaker meaning can't just be a matter of intending to cause a certain belief — it must be intending to cause a certain belief in a certain way. But what, in addition to intending to cause the belief, is required for meaning that p ? Grice's idea was that one must not only intend to cause the audience to form a belief, but also intend that they do so on the basis of their recognition of the speaker's intention. This condition is not met in the above example: I don't expect you to believe that I have uttered a certain sentence on the basis of your recognition of my intention that you do so; after all, you'd believe this whether or not I wanted you to. This is all to the good.

However, even if [G] can be given a fairly plausible motivation, and fits many cases rather well, it is also open to some convincing counterexamples. Three such types of cases are: (i) cases in which the speaker means p by an utterance despite knowing that the audience already believes p , as in cases of reminding or confession; (ii) cases in which a speaker means p by an utterance, such as the conclusion of an argument, which the speaker intends an audience to believe on the basis of evidence rather than recognition of speaker intention; and (iii) cases in which there is no intended audience at all, as in uses of language in thought. These cases call into question whether there is any connection between speaker-meaning and intended effects stable enough to ground an analysis of the sort that Grice envisaged; it is still a matter of much controversy whether an explanation of speaker meaning descended from [G] can succeed. Despite this controversy, the Gricean analysis is probably still the closest thing to orthodoxy when it comes to foundational theories of meaning. (For developments of the Gricean program, see — in addition to the classic essays in Grice (1989) — Schiffer (1972); Neale (1992); Davis (2002). For an extended criticism, see Schiffer (1987).)

3.1.2 *Meaning, belief, and convention*

An important alternative to the Gricean analysis, which shares the Gricean's commitment to a mentalist analysis of meaning in terms of the contents of mental states, is the analysis of meaning in terms of the *beliefs* rather than the intentions of speakers.

It is intuitively plausible that such an analysis should be possible. After all, there clearly are regularities which connect utterances and the beliefs of speakers; roughly, it seems that, for the most part, speakers seriously utter a sentence which (in the context) means p only if they also believe p . One might then, try to analyze meaning directly in terms of the beliefs of language users, by saying that what it is for a sentence S to express some proposition p is for it to be the case that, typically, members of the community would not utter S unless they believed p . However, we can imagine a community in which there is some action which everyone would only perform were they to believe some proposition p , but which is such that no member of the community knows that

any other member of the community acts according to a rule of this sort. It is plausible that in such a community, the action-type in question would not express the proposition p , or indeed have any meaning at all.

Because of cases like this, it seems that regularities in meaning and belief are not sufficient to ground an analysis of meaning. For this reason, many proponents of a mentalist analysis of meaning in terms of belief have sought instead to analyze meaning in terms of *conventions* governing such regularities. There are different analyses of what it takes for a regularity to hold by convention (see CONVENTION); according to one important view, a sentence S expresses the proposition p if and only if the following three conditions are satisfied: (1) speakers typically utter S only if they believe p and typically come to believe p upon hearing S , (2) members of the community believe that (1) is true, and (3) the fact that members of the community believe that (1) is true, and believe that other members of the community believe that (1) is true, gives them a good reason to go on acting so as to make (1) true. (This a simplified version of the theory defended in Lewis (1975).) For critical discussion of this sort of analysis of meaning, see Burge (1975), Hawthorne (1990), Laurence (1996), and Schiffer (2006).

3.1.3 *Mental representation-based theories*

The two sorts of mentalist theories sketched above both try to explain meaning in terms of the relationship between linguistic expressions and propositional attitudes of users of the relevant language. But this is not the only sort of theory available to a theorist who wants to analyze meaning in terms of mental representation. A common view in the philosophy of mind and cognitive science is that the propositional attitudes of subjects are underwritten by an internal language of thought, comprised of mental representations. (See THE COMPUTATIONAL THEORY OF MIND.) One might try to explain linguistic meaning directly in terms of the contents of mental representations, perhaps by thinking of language processing as pairing linguistic expressions with mental representations; one could then think of the meaning of the relevant expression for that individual as being inherited from the content of the mental representation with which it is paired. For discussion of this sort of theory, see Laurence (1996).

Just as proponents of Gricean and convention-based theories typically view their theories as only the first stage in an analysis of meaning — because they analyze meaning in terms of another sort of mental representation — so proponents of mental representation-based theories will typically seek to provide an independent analysis of contents of mental representations. For a good overview of attempts to provide the latter sort of theory, see the essays in Stich and Warfield (1994).

3.2 *Non-mentalist theories*

As noted above, not all foundational theories of meaning attempt to explain meaning in terms of mental representation. One might be inclined to pursue a

non-mentalist foundational theory of meaning for a number of reasons; for example, one might be skeptical about the mentalist theories on offer; one might think that mental representation should be analyzed in terms of linguistic representation, rather than the other way around; or one might think that representation should be analyzable in non-representational terms, and doubt whether there is any true explanation of mental representation suitable to accompany a mentalist reduction of meaning to mental representation.

All non-mentalist foundational theories of meaning attempt to explain the meanings of expressions in terms of their use. This is not to say very much; one might say the same about mentalist theories. (Gricean theories, for example, say that what counts is using the expression with a certain communicative intention.) To give a non-mentalist foundational theory of meaning is to say which *aspects* of the use of an expression determine meaning — and do so without making use of facts about which propositional attitudes and mental representations accompany that use. In what follows I'll briefly discuss some of the aspects of the use of expressions which proponents of non-mentalist theories have taken to explain their meanings.

3.2.1 Causal origin

In *Naming and Necessity*, Kripke suggested that the reference of a name could be explained in terms of the history of use of that name, rather than by descriptions associated with that name by its users. In the standard case, Kripke thought, the right explanation of the reference of a name could be divided into an explanation of the name's *introduction* as name for this or that — an event of 'baptism' — and its successful *transmission* from one speaker to another.

One approach to the theory of meaning is to extend Kripke's remarks in two ways: first, by suggesting that they might serve as an account of meaning, as well as reference;⁷ and second, by extending them to parts of speech other than names. (See, for discussion, Devitt (1981).) In this way, we might aim to explain the meanings of expressions in terms of their causal origin.

One point worth noting about this sort of theory is that it is far from clear that it is really a non-mentalist theory. One might think that introducing a term involves *intending* that it stand for some object or property, and one might think that transmission of a term from one speaker to another involves the latter *intending* to use it in the same way as the former. If so, then perhaps causal theories, no less than Gricean theories, analyze meaning in terms of the intentions of language users.

There are two standard problems for causal theories of this sort (whether or not they are elaborated in a mentalist or a non-mentalist way). The first is the problem of extending the theory from the case of names to other sorts of vocabulary for which the theory seems less natural. Examples which have seemed to many to be problematic are empty names and non-referring theoretical terms, logical vocabulary, and predicates which, because their content does not seem

⁷This is most natural for views, like Millian-Russellian views, which make meaning closely related to the entity in the world for which the word stands.

closely related to the properties represented in perceptual experience, are not intuitively linked to any initial act of ‘baptism.’ The second problem, which is sometimes called the ‘*qua* problem,’ is the problem of explaining which of the many causes of a term’s introduction should determine its content. Suppose that the term ‘water’ was introduced in the presence of a body of H₂O. What made it a term for this substance, rather than for liquid in general, or colorless liquid, or colorless liquid in the region of the term introduction? The proponent of a causal theory owes some answer to this question; see for discussion Devitt and Sterelny (1987).

3.2.2 *Truth-maximization and the principle of charity*

Causal theories aim to explain meaning in terms of the relations between expressions and the objects and properties they represent. A very different sort of foundational theory of meaning which maintains this emphasis on the relations between expressions and the world gives a central role to a *principle of charity* which holds that (modulo some qualifications) the right assignment of meanings to the expression of a subject’s language is that assignment of meanings which maximizes the truth of the subject’s utterances.

An influential proponent of this sort of view was Donald Davidson, who stated the motivation for the view as follows:

“A central source of trouble is the way beliefs and meanings conspire to account for utterances. A speaker who holds a sentence to be true on an occasion does so in part because of what he means, or would mean, by an utterance of that sentence, and in part because of what he believes. If all we have to go on is the fact of honest utterance, we cannot infer the belief without knowing the meaning, and have no chance of inferring the meaning without the belief.” (Davidson (1973a), 314; see also Davidson (1973b)) .

Davidson’s idea was that attempts to state the facts in virtue of which expressions have a certain meaning for a subject face a kind of dilemma: if we had an independent account of what it is for an agent to have a belief with a certain content, we could ascend from there to an account of what it is for a sentence to have a meaning; if we had an independent account of what it is for a sentence to have a meaning, we could ascend from there to an account of what it is for an agent to have a belief with a certain content; but in fact neither sort of independent account is available, because many assignments of beliefs and meanings are consistent with the subject’s linguistic behavior. Davidson’s solution to this dilemma is that we must define belief and meaning together, in terms of an independent third fact: the fact that the beliefs of an agent, and the meanings of her words, are whatever they must be in order to maximize the truth of her beliefs and utterances.

By tying meaning and belief to truth, this sort of foundational theory of meaning implies that it is impossible for anyone who speaks a meaningful language to be radically mistaken about the nature of the world; and this implies

that certain levels of radical disagreement between a pair of speakers or communities will also be impossible (since the beliefs of each community must be, by and large, true). This is a consequence of the view which Davidson embraced (see Davidson (1974)); but one might also reasonably think that radical disagreement, as well as radical error, are possible, and hence that any theory, like Davidson's, which implies that they are impossible must be mistaken (see Chapter 13 of Soames (2003) for discussion).

A different sort of worry about a theory of this sort is that the requirement that we maximize the truth of the utterances of subjects hardly seems sufficient to determine the meanings of the expressions of their language. It seems plausible, offhand, that there will be many different interpretations of a subject's language which will be tied on the measure of truth-maximization; one way to see the force of this sort of worry is to recall the point, familiar from our discussion of possible worlds semantics in §2.1.5 above, that a pair of sentences can be true in exactly the same circumstances and yet differ in meaning. One worry is thus that a theory of Davidson's sort will entail an implausible indeterminacy of meaning. For Davidson's fullest attempt to answer this sort of worry, see Chapter 3 of Davidson (2005).

3.2.3 *Regularities in use*

A different way to develop a non-mentalist foundational theory of meaning focuses less on relations between subsentential expressions or sentences and bits of non-linguistic reality and more on the regularities which govern our use of language. Views of this sort have been defended by a number of authors; in what follows I sketch the version of the view developed in Horwich (1998, 2005).

Horwich's core idea is that our acceptance of sentences is governed by certain laws, and, in the case of non-ambiguous expressions, there is a single 'acceptance regularity' which explains all of our uses of the expression. The type of acceptance regularity which is relevant will vary depending on the sort of expression whose meaning is being explained. For example, our use of a perceptual term like 'red' might be best explained by the following acceptance regularity:

The disposition to accept 'that is red' in response to the sort of visual experience normally provoked by a red surface.

whereas, in the case of a logical term like 'and,' the acceptance regularity will involve dispositions to accept inferences involving pairs of sentences rather than dispositions to respond to particular sorts of experiences:

The disposition to accept the two-way argument schema " $p, q // p$ and q ."

As these examples illustrate, it is plausible to think that a strength of a view like Horwich's is its ability to handle expressions of different categories.

Like its companions, Horwich's theory is also open to some objections. One might worry that his use of the sentential attitude of acceptance entails a lapse

into mentalism, if acceptance either just is, or is analyzed in terms of, beliefs. There is also a worry — which affects other ‘use’ or ‘conceptual role’ or ‘functional role’ theories of meaning — that Horwich’s account implies the existence of differences in meaning which do not exist; it seems, for example, that two people’s use of some term might be explained by distinct basic acceptance regularities without their meaning different things by that term. Schiffer (2000) discusses the example of ‘dog’, and the differences between the basic acceptance regularities which govern the use of the term for the blind, the biologically unsophisticated, and people acquainted only with certain sorts of dogs.⁸

3.2.4 Social norms

This last concern about Horwich’s theory stems from the fact that the theory is, at its core, an individualist theory: it explains the meaning of an expression for an individual in terms of properties of that individual’s use of the term. A quite different sort of use theory of meaning turns from the laws which explain an individual’s use of a word to the norms which, in a society, govern the use of the relevant terms. Like the other views discussed here, the view that meaning is a product of social norms of this sort has a long history; it is particularly associated with the work of the later Wittgenstein and his philosophical descendants. (See especially Wittgenstein (1953).)

An important recent defender of this sort of view is Robert Brandom. On Brandom’s view, a sentence’s meaning is due to the conditions, in a given society, under which it is correct or appropriate to perform various speech acts involving the sentence. To develop a theory of this sort, one must do two things. First, one must show how the meanings of expressions can be explained in terms of these normative statuses — in Brandom’s (slightly nonstandard) terms, one must show how semantics can be explained in terms of pragmatics. Second, one must explain how these normative statuses can be instituted by social practices. For details, see Brandom (1994), in which the view is developed at great length; for a critical discussion of Brandom’s attempt to carry out the second task above, see Rosen (1997).

References

- Michael Beaney (editor) 1997. *The Frege Reader*. Oxford: Basil Blackwell.
- Robert Brandom, 1994. *Making It Explicit*. Cambridge, MA: Harvard University Press.
- David Braun, 1993. Empty Names. *Noûs* 27(4):449–469.

⁸Horwich (1998) (85-6) suggests that cases of this sort can be handled by adding to his account a clause which makes it sufficient to mean (for example) DOG by ‘dog’ that one be disposed to defer to experts whose use of the term *is* explained by the canonical basic acceptance regularity, so long as one’s use is ‘close enough’ to theirs. But it is hard to see why the absence of such a disposition, perhaps out of simple obstinacy or distaste for experts, should disqualify a biologically unsophisticated person from meaning by ‘dog’ what we do.

- David Braun and Jennifer Saul, 2002. Simple Sentences, Substitutions, and Mistaken Evaluations. *Philosophical Studies* 111:1–41.
- Tyler Burge, 1975. On Knowledge and Convention. *Philosophical Review* 84:2:249–255.
- Tyler Burge, 1986. On Davidson’s ‘Saying That’. In *Truth and Interpretation: Perspectives on the Philosophy of Donald Davidson*, 190–210. Oxford: Blackwell Publishing.
- Ben Caplan, 2005. Against Widescopism. *Philosophical Studies* 125(2):167–190.
- Herman Cappelen and Ernie Lepore, 2005. *Insensitive Semantics: A Defence of Semantic Minimalism and Speech Act Pluralism*. Blackwell Publishing.
- Rudolf Carnap, 1947. *Meaning and Necessity: A Study in Semantics and Modal Logic*. University of Chicago Press.
- David Chalmers, 2004. Epistemic Two-Dimensional Semantics. *Philosophical Studies* 118(1-2):153–226.
- David Chalmers, 2006. The Foundations of Two-Dimensional Semantics. In *Two-Dimensional Semantics*, edited by Manuel García-Carpintero and Josep Macià, 55–140. New York: Oxford University Press.
- David Chalmers, 2007. Propositions and Attitude Ascriptions: A Fregean Account. *ms* .
- Noam Chomsky, 2000. *New Horizons in the Study of Language and Mind*. Cambridge: Cambridge University Press.
- Donald Davidson, 1967/1984. Truth and Meaning. In Davidson (1984), 17–36.
- Donald Davidson, 1968. On Saying That. *Synthese* 19:130–146.
- Donald Davidson, 1973a. Belief and the Basis of Meaning. *Synthese* 27:313–328.
- Donald Davidson, 1973b. Radical Interpretation. *Dialectica* 27:313–328.
- Donald Davidson, 1974. On the Very Idea of a Conceptual Scheme. In Davidson (1984), 183–198.
- Donald Davidson, 1976. Reply to Foster. In Evans and McDowell (1976), 33–41.
- Donald Davidson, 1984. *Inquiries into Truth and Interpretation*. Oxford: Oxford University Press.
- Donald Davidson, 2005. *Truth and Predication*. Cambridge, MA: Harvard University Press.
- Wayne Davis, 2002. *Meaning, Expression, and Thought*. New York: Cambridge University Press.

- Michael Devitt, 1981. *Designation*. New York: Columbia University Press.
- Michael Devitt and Kim Sterelny, 1987. *Language and Reality: An Introduction to the Philosophy of Language*. Cambridge, MA: MIT Press.
- Michael Dummett, 1981. *The Interpretation of Frege's Philosophy*. Harvard University Press.
- Gareth Evans, 1981. Understanding Demonstratives. In his *Collected Papers*, 291–321. Oxford: Oxford University Press.
- Gareth Evans and John McDowell (editors) 1976. *Truth and Meaning: Essays in Semantics*. Oxford: Oxford University Press.
- Kit Fine, 2007. *Semantic Relationism*. Blackwell Publishing.
- John Foster, 1976. Meaning and Truth Theory. In Evans and McDowell (1976), 1–32.
- Gottlob Frege, 1879/1997. Begriffsschrift. In Beaney (1997), 47–79.
- Gottlob Frege, 1892/1960. On Sense and Reference. In *Translations from the Philosophical Writings of Gottlob Frege*, edited by Peter Geach and Max Black, 56–78. Oxford: Basil Blackwell.
- Gottlob Frege, 1906/1997. A Brief Survey of My Logical Doctrines. In Beaney (1997), 299–300.
- Paul Grice, 1957. Meaning. In Grice (1989), 213–223.
- Paul Grice, 1969. Utterer's Meaning and Intentions. In Grice (1989), 86–116.
- Paul Grice, 1989. *Studies in the Way of Words*. Cambridge, MA: Harvard U.P.
- John Hawthorne, 1990. A Note on 'Languages and Language'. *Australasian Journal of Philosophy* 68:1:116–118.
- Irene Heim and Angelika Kratzer, 1998. *Semantics in Generative Grammar*. Oxford: Blackwell Publishing.
- Paul Horwich, 1998. *Meaning*. Oxford: Clarendon Press.
- Paul Horwich, 2005. *Reflections on Meaning*. Oxford: Clarendon Press.
- Mark Johnston, 1988. The End of the Theory of Meaning. *Mind & Language* 3:1:28–42.
- Hans Kamp, 1971. Formal Properties of 'Now'. *Theoria* 37(3):227–273.
- David Kaplan, 1989. Demonstratives. In *Themes from Kaplan*, edited by Joseph Almog, John Perry, and Howard Wettstein, 481–563. Oxford: Oxford University Press.

- Max Kölbel, 2001. Two Dogmas of Davidsonian Semantics. *Journal of Philosophy* 98:12:613–635.
- Saul Kripke, 1972. *Naming and Necessity*. Cambridge, MA: Harvard University Press.
- Saul Kripke, 1979. A Puzzle About Belief. In *Meaning and Use*, edited by Avishai Margalit, 239–283. Boston: D. Reidel Publishing Company.
- Saul Kripke, 1982. *Wittgenstein on Rules and Private Language: An Elementary Exposition*. Cambridge, MA: Harvard U.P.
- Richard Larson and Peter Ludlow, 1993. Interpreted Logical Forms. *Synthese* 95:305–355.
- Richard Larson and Gabriel Segal, 1995. *Knowledge of Meaning*. Cambridge, MA: MIT Press.
- Stephen Laurence, 1996. A Chomskian Alternative to Convention-Based Semantics. *Mind* 105:269–301.
- Ernest LePore and Barry Loewer, 1989. You Can Say That Again. *Midwest Studies in Philosophy* 14:338–356.
- David Lewis, 1970. General Semantics. *Synthese* 22:18–67.
- David Lewis, 1975. Languages and Language. 3–35. Reprinted in Lewis (1983), 163–188.
- David Lewis, 1980. Index, Context, and Content. In *Philosophy and Grammar*, edited by Stig Kanger and Sven Ohman. Reidel.
- David Lewis, 1983. *Philosophical Papers*, volume 1. Oxford: Oxford U.P.
- John McDowell, 1977. On the Sense and Reference of a Proper Name. *Mind* 86:159–185.
- James McGilvray, 1998. Meanings Are Syntactically Individuated and Found in the Head. *Mind and Language* 13(225–280).
- Jonathan McKeown-Green, 2002. *The Primacy of Public Language*. PhD Dissertation, Princeton University: unpublished.
- Stephen Neale, 1992. Paul Grice and the Philosophy of Language. *Linguistics & Philosophy* 15:5:509–559.
- Paul Pietroski, 2003. The Character of Natural Language Semantics. In *Epistemology of Language*. Oxford: Oxford University Press.
- Paul Pietroski, 2005. Meaning Before Truth. In *Contextualism in Philosophy*. Oxford: Oxford University Press.

- Alvin Plantinga, 1978. The Boethian Compromise. *American Philosophical Quarterly* 15:129–138.
- W. V. Quine, 1960. *Word and Object*. Cambridge, MA: MIT Press.
- W. V. Quine, 1970. *Philosophy of Logic*. New Jersey: Prentice Hall.
- Mark Richard, 1981. Temporalism and Eternalism. *Philosophical Studies* 39(1):1–13.
- Gideon Rosen, 1997. Who Makes the Rules around Here? *Philosophy and Phenomenological Research* 57(1):163–171.
- Bertrand Russell, 1903. *The Principles of Mathematics*. Cambridge University Press.
- Nathan Salmon, 1986. *Frege's Puzzle*. Atascadero, CA: Ridgeview Publishing Company.
- Nathan Salmon, 1990. A Millian Heir Rejects the Wages of *Sinn*. In *Propositional Attitudes: The Role of Content in Logic, Language, and Mind*, edited by C. Anthony Anderson and Joseph Owens, 117–139. Stanford, CA: CSLI Press.
- Stephen Schiffer, 1972. *Meaning*. Oxford: Oxford University Press.
- Stephen Schiffer, 1987. *Remnants of Meaning*. Cambridge, MA: MIT Press.
- Stephen Schiffer, 2000. Horwich on Meaning. *Philosophical Quarterly* 50(201):527–536.
- Stephen Schiffer, 2006. Two Perspectives on Knowledge of Language. *Philosophical Issues* 16:275–287.
- Scott Soames, 1988. Direct Reference, Propositional Attitudes, and Semantic Content. In *Propositions and Attitudes*, edited by Nathan Salmon and Scott Soames, 197–239. Oxford: Oxford University Press.
- Scott Soames, 1992. Truth, Meaning, and Understanding. *Philosophical Studies* 65:17–35.
- Scott Soames, 1998. The Modal Argument: Wide Scope and Rigidified Descriptions. *Noûs* 32(1):1–22.
- Scott Soames, 1999. Skepticism About Meaning: Indeterminacy, Normativity, and the Rule-following Paradox. *Canadian Journal of Philosophy* 23:211–249.
- Scott Soames, 2002. *Beyond Rigidity: The Unfinished Semantic Agenda of Naming and Necessity*. Oxford: Oxford University Press.
- Scott Soames, 2003. *Philosophical Analysis in the Twentieth Century: Volume 2, the Age of Meaning*. Princeton, NJ: Princeton University Press.

- David Sosa, 2001. Rigidity in the Scope of Russell's Theory. *Noûs* 35(1):1–38.
- Jeff Speaks, 2006. Truth Theories, Translation Manuals, and Theories of Meaning. *Linguistics & Philosophy* 29(4):487–505.
- Robert Stalnaker, 1984. *Inquiry*. Cambridge, MA: MIT Press.
- Stephen Stich and Ted Warfield (editors) 1994. *Mental Representation: A Reader*. Cambridge, MA: Basil Blackwell.
- Alfred Tarski, 1936. The Semantic Conception of Truth and the Foundations of Semantics. *Philosophy and Phenomenological Research* 4:3:341–376.
- Ludwig Wittgenstein, 1953. *Philosophical Investigations*. Translated by G. E. M. Anscombe. New York: MacMillan, 3 edition.