

# Frege's theory of reference

Jeff Speaks  
August 26, 2011

1. Two approaches to logic .....	1
1.1. The connection between logic and the philosophy of language	
1.2. Subject/predicate and the theory of the syllogism	
1.3. The problem of multiple generality	
1.4. Frege's function/argument analysis	
2. Reference and truth-value.....	4
2.1. Bedeutung	
2.2. The reference of names and predicates	
2.3. Quantification and generality	

Much of analytic philosophy is traceable, in part, to Frege's philosophy of language. One good way to understand Frege's philosophy of language is to begin with Frege's logic, and see how this led to his views about reference; and then to see why, ultimately, Frege came to think that his theory of reference needed to be supplemented with a theory of sense.

## 1. TWO APPROACHES TO LOGIC

### *1.1. The connection between logic and the philosophy of language*

Let's begin by asking: What does logic have to do with the philosophy of language? One of the tasks of the philosophy of language is to explain the relations between linguistic expressions and the world in virtue of which certain sentences are true or false. Now, it seems plausible that any account of what makes certain sentences true or false will be given in terms of the connections between various sub-sentential expressions and objects and properties in the world. The fact that, e.g.,

Grass is green.

is true evidently has something to do with the fact that 'grass' is a word for the plants found frequently in front yards and that 'green' is a word for a particular color. This indicates that any complete philosophy of language will have to provide some sort of analysis of sentences into their parts which can explain how the conditions which would make those sentences true are determined by those parts.

Deductive logic is the study of the truth-involving relations between sentences; logicians are especially interested in giving a general answer to the question: When does the truth of one set of sentences guarantee the truth of some other sentence? (I.e.: What does it take for one sentence to follow from some others?) It's clear that giving a general answer

to this question will also involve giving some kind of analysis of sentences into their parts, since whether one sentence follows from another is typically a matter of relations between parts of those sentences.

The important point to note is that the two kinds of analyses of sentences are likely to coincide. The latter kind of analysis is meant to explain the truth-involving relations between sentences. The most natural way to pursue this kind of analysis would be to explain the structural features of sentences in virtue of which they have certain truth conditions, and then to explain logical consequence in terms of these features. But, if this is so, then the most natural way to give a theory of logical consequence for a language will also give us the kind of structural explanation of language-world relations which we are looking for in the philosophy of language.

### 1.2. *Subject/predicate and the theory of the syllogism*

To understand why Frege's analysis of language was revolutionary, it is useful to have some grasp of the subject/predicate analysis of language which it overthrew; the easiest way to approach this is via Aristotle's theory of the syllogism.

Aristotle's *Prior Analytics* is devoted to explaining the nature of demonstration; he introduces the topic by saying

“We must ...define a premiss, a term, and a syllogism ...and after that, the inclusion or non-inclusion of one term in another as in a whole, and what we mean by predicating one term of all, or none, of another.” (24a10-15)

Premises are sentences which, Aristotle says, will be

“an affirmation or denial of something concerning something else.” (24a26-27)

He goes on to distinguish three different ways in which something may be said of something else (modes of attribution): universal, particular, and indefinite (25a2-5). This certainly seems to assume that the relevant logical features of a sentence are determined by the relationship between the predicate — what is said of — and the subject — the thing of which it is said.

When he goes on to discuss demonstration and to give an account of (what we might call) the types of valid argument, he does so in terms of this catalogue of distinctions between ways in which something is said of something else. The following syllogism is an example:

[Major Premise] All mammals are animals.  
[Minor Premise] All dogs are mammals.  
[Conclusion] All dogs are animals.

This is the ‘first figure’ of the syllogism, in which the subject of the major premise is the predicate of the minor premise, the predicates of the major premise and the conclusion are the same, and the subjects of the minor premise and the conclusion are the same. Each of the premises and the conclusion of this instance of the first figure are both universal predications (as opposed to particular or indefinite ones) and are affirmations (as opposed to negations/denials). The first figure, with all sentences universal and affirmative, is valid. (Some changes would make it invalid, as if the major premise were particular rather than universal. Other changes would preserve its validity, as if the minor premise and conclusion were both particular rather than universal.)

The central point to notice: the most general distinctions between classes of sentences are distinctions between modes of predication. This seems to presuppose that we can explain valid inference by suitably categorizing the subject, and the predicate, of the sentences involved in the relevant argument.

### *1.3. The problem of multiple generality*

One persistent problem for the subject/predicate analysis, and hence for Aristotelian logic, was posed by sentences of multiple generality — i.e., sentences which contain more than one expression of generality, such as ‘someone’ or ‘everything’, like

Everyone loves someone.

The problem of multiple generality is the problem of accounting for the logical relations in which sentences of multiple generality stand. An example of the kind of logical relation which proved problematic for Aristotelian logic and its successors was the fact that

Someone is loved by everyone.

seems to logically entail

Everyone loves at least one person.

It is difficult to see how to give a systematic account of entailments of this sort in Aristotelian logic.

### *1.4. Frege’s function/argument analysis*

Frege rejected the Aristotelian subject/predicate analysis of sentences, and thought that it went wrong by following the superficial appearance of natural language too closely. As he said, “A distinction of subject and predicate finds no place in my way of representing a judgement.” (*Begriffsschrift* §3)

In place of Aristotle's analysis of sentences into subject and predicate, Frege proposed an analysis of sentences into one or more *functions* and one or more *arguments*.

A familiar example of an expression which stands for a function is '+' This sign stands for a function from pairs of numbers to numbers; i.e., the addition function takes pairs of numbers as its arguments and yields a number as its value. To understand how expressions of natural language might be akin to '+', we should ask a number of questions: (i) Which expressions stand for functions, and which for arguments to those functions? (ii) What kinds of things do these functions have as their arguments and values?

To answer these questions, we turn to Frege's theory of *Bedeutung*.

## 2. REFERENCE AND TRUTH-VALUE

### 2.1. *Bedeutung*

Remember that our aim is to give a logic — i.e., a theory of validity — and that our idea is that we can do so by giving a theory of the way in which the parts of sentences combine to determine that sentence's truth-value.

*Bedeutung*, now standardly translated as "reference", was Frege's name for that property of a linguistic expression which is that expression's contribution to determining the truth or falsity of sentences in which it occurs. So it is natural to think of the reference of an expression — that thing Frege was trying to give an account of — to, as a first approximation, *power to affect truth-value*. (Frege expressed this by saying that the reference of a sentence is a certain special sort of object — a truth-value — but this is not essential to the view.)

If the reference of an expression is that expression's power to affect the truth-value of sentences in which it occurs, and if a theory of reference is to give us a satisfying explanation of the relations between sentences and the world in virtue of which some are true and some false, then there had better be a systematic connection between the references of sub-sentential expressions and the truth-values of sentences. In particular, the following connection between reference and truth had better be true:

The truth-value of a sentence is a function of the references of the expressions which compose the sentences, along with the way in which they are combined (and along with, of course, the relevant facts about the world).

E.g., the truth or falsity of

John loves Mary.

should be determined by the references of ‘John’, ‘Mary’, ‘loves’, the way the three terms are combined, and the relevant extra-linguistic facts about John and Mary.

This gives us an initial test for when two expressions have the same, or different, reference: two expressions have the same reference iff substitution of one for another in a sentence never changes that sentence’s truth value.

## 2.2. *The reference of names and predicates*

There is a sense in which, for Frege, the foundations of the theory of reference are to be found in the theory of the references of proper names, and the view that the reference of a proper name is an object.

Names, for Frege, were a broad category; they included not just proper names, like “Jeff Speaks”, but also definite descriptions, like “the teacher of PHIL 83104.” The reference of all such expressions is the object singled out by the name.

What is the reference of a predicate, like “is a teacher”? Well, we know (i) that reference is power to affect truth-value, (ii) the reference of a name is an object, and (iii) predicates are expressions which combine with names to form sentences, which are true or false. So the reference of a predicate is, to put the point metaphorically, what is left over from a truth-value when you subtract an object — or, less metaphorically something which combines with an object to determine a truth-value. A natural candidate for the reference of a monadic predicate is thus: *a function from objects to truth-values*.

How would you extend this idea to give a theory of the reference of relational predicates? How about adverbs, like “quickly”?

Frege thought that sentences involving proper names with no reference must lack a truth-value, on the grounds that if an expression lacks a reference, then all complex expressions containing it must lack a reference. While this principle seems plausible for some cases (“This is a nice red one”) it seems much less plausible in other cases (“The largest prime number is less than 100”). But endorsing a broadly Fregean approach to the theory of reference doesn’t force us into Frege’s particular views about the truth-values of sentences containing empty names.

## 2.3. *Quantification and generality*

One of the principal advantages of this sort of analysis of language over its Aristotelian predecessor arises from its treatment of expressions of generality, like “everyone”, “something”, etc.

Frege analyzed these expressions of generality as *quantifiers*, which we now express using the symbols  $\exists$  and  $\forall$ . (or now, we will just focus on these two — the existential and universal quantifiers.) These are expressions which combine with variables to form expressions which can in turn combine with monadic predicates to form sentences.

Assuming that, as above, we already have a handle on our theory of reference for monadic predications, we can explain what it takes for sentences involving these quantifiers to be true as follows:

$\exists x Fx$  iff for some object  $a$ ,  $Fa$

$\forall x Fx$  iff for every object  $a$ ,  $Fa$

(Can either of these quantifiers be defined in terms of the other?)

Given this, how should we think of the reference of  $\exists$  and  $\forall$ ? Well, we know that, like names, they combine with monadic predicates to form sentences. But their reference can't be an object — what would be the object for which “something” stands? Instead, it is natural to take them to be functions from the references of monadic predicates to truth-values — i.e., functions from functions from objects to truth-values to truth-values.

Thought of in this way, the reference of  $\exists$  is the following second-order function: the function which (i) has value “true” when given as argument a function from objects to truth-values which has value “true” for at least one object, and (ii) has value “false” when given as argument any other function from objects to truth-values.

Consider how this might help with the problem of multiple generality, specifically with the problem of explaining the validity of the inference from

[1] Someone is loved by everyone.

to

[2] Everyone loves at least one person.

Well, let's think about the analysis of these sentences which Frege's ideas gives us. The form of [1] might be represented as

$\exists x \forall y (y \text{ loves } x)$

But how are we supposed to understand sentences which have two quantifiers lined up in front? We think of the sentence as built up in stages. First, we have the two-place predicate “loves.” We then combine this with the universal quantifier to form the one place predicate

$\forall y (y \text{ loves } \_ \_)$

This predicate is true of an object  $o$  iff  $o$  is such that everyone loves it.<sup>1</sup> If we combine this predicate with the existential quantifier, we get a sentence which is true if and only if there is some object which is such that everyone loves it.

Using parallel reasoning, we can think of the form of [2] as

$\forall x \exists y (x \text{ loves } y)$

which is true iff, roughly, the property of loving someone is instantiated by everything. (Here I'm just simplifying by assuming that we are just quantifying over people.)

Using these analyses, and the rules of inference for  $\exists$  and  $\forall$ , it is easy to prove that [1] implies [2]. Frege's treatment of generality, embedded in his function/argument analysis of the form of sentences, thus solves a very general problem which the subject/predicate analysis could not.

.....

One can imagine trying to expand the above sketch into a full theory of reference for a natural language. This would obviously be a highly non-trivial undertaking. But we can also ask: in principle, if we could give such an account, would this be a satisfactory account of the contents, or meanings, of expressions of the language? Frege came to think not; his reasons for this are our next topic.

---

<sup>1</sup> At this stage, you might object: "Wait a minute; I thought that the reference of "everyone" was a function from functions from objects to truth-values to truth-values; how can it combine with a two-place predicate — whose reference is a function from ordered pairs to truth-values — to form a monadic predicate — whose reference is a function from objects to truth-values? Does this mean that "everyone" is ambiguous?" This is an excellent question, and shows one way in which the sketch of the theory of reference of quantifiers which I am giving here is an oversimplification.