

Føllesdal and Frege on Reference

Øystein Linnebo*

0. Introduction

In a recent essay on reference, Dagfinn Føllesdal writes the following.

One must distinguish between on the one hand a *two-sorted semantics*, where terms that refer are treated quite differently from general terms and other non-referring expressions, and on the other hand a *theory of reference*, that is an account of the relation between referring expressions and their objects. What I argued for in my dissertation was the former. I recognized that talk about modality, knowledge, belief, causation, change, probability, ethics, etc., makes sense only if the referring expressions in our language have a semantics that is very different from that of general terms. At that time I thought that the referring expressions actually succeed in relating to their objects regardless of how the world and our theories about the world change, and I was at a loss as to how to explain how this could happen. (Føllesdal 1997, p. 359; see also Føllesdal 2004, p. xxviii.)

In this important passage Føllesdal distinguishes between *semantics proper* and *the theory of reference*. He explains how in his dissertation he developed a novel “two-sorted semantics” where singular terms are treated very differently from general terms: where a general term can apply to different objects in different possible worlds, a singular term applies to the same object in all possible worlds (at least in which the object exists). But Føllesdal also admits that early in his career he didn’t know how to develop a theory of reference, that is, an account of the relation that obtains between a referring expression and its referent.

In this essay I first identify some desiderata for a theory of reference that appear in Føllesdal’s later work (Section 1). We will see that these desiderata are highly Fregean in character. Next I outline a Frege-inspired theory of (a certain core form of) reference that I have been developing in my own

* Birkbeck, University of London

work (Section 2). Finally I argue that this theory (as far as it goes) satisfies Føllesdal's desiderata (Section 3). In particular, I argue that, when the distinction between semantics proper and the theory of reference is carefully heeded, my Frege-inspired theory of reference can be seen to be fully compatible with Føllesdal's "two-sorted semantics." The Fregeanism that I advocate is thus highly selective: its only direct concern is the theory of reference, not semantics proper.

1. Føllesdal on Singular Terms

In his 1961 dissertation, written at Harvard under the supervision of W.V. Quine, Føllesdal investigates Quine's famous objection to quantified modal logic. A careful examination of Quine's objection leads Føllesdal to formulate the notion of what we now know as a *rigid designator* (which Føllesdal calls a *genuine singular term*). A rigid designator is, as we all know, a term that denotes the same object in all possible worlds (or at least in all worlds in which this object exists). Føllesdal shows us how Quine's objection can be avoided by carefully distinguishing between rigid and non-rigid designators and by holding that singular terms, unlike definite descriptions, are rigid designators. This is an important contribution to semantics in which Føllesdal articulates and defends one of the most important doctrines now usually associated with Saul Kripke.

To understand the similarities and differences between Føllesdal and Kripke, it will be useful to distinguish between what I will call *semantics* and *meta-semantics*. This distinction generalizes and subsumes the distinction Føllesdal draws in the passage just quoted between a semantics and a theory of reference.¹ Semantics typically takes the form of a theory of *semantic values*, where the semantic value [E] of an expression E is the contribution that E makes to the truth-values of sentences in which it occurs.² Following Frege, it is argued that semantic values are subject to a *principle of compositionality*, according to which the semantic value of a complex expression

¹ My distinction between semantics and meta-semantics is thus the same as Stalnaker's distinction between "descriptive" and "foundational" semantics. See e.g. Stalnaker 1997. The same distinction is found in the work of various other philosophers as well, for instance Michael Dummett, Richard Heck, and Jason Stanley.

² I will use boldface for all meta-linguistic variables.

is determined as a function of the semantic values of its individual sub-expressions. For instance, the semantic value of an atomic sentence $\mathbf{P}(\mathbf{a}_1, \dots, \mathbf{a}_n)$ is functionally determined as $[\mathbf{P}]([\mathbf{a}_1], \dots, [\mathbf{a}_n])$. Following Frege again, the semantic value of a sentence is often taken to be just its truth-value, and the semantic value of a proper name, its referent. If these two assumptions are accepted, then the principle of compositionality will allow us to determine the kinds of semantic values had by other types of expressions; for instance, the semantic value of a one-place predicate must be a function from objects to truth-values.

I said above that Føllesdal's discovery of the notion of a rigid designator was an important contribution to semantics. When I said this, I had in mind semantics in the precise sense that I have just explained. The semantic value of a rigid designator is an object. And this semantic value remains associated with the term throughout all possible worlds. This contrasts with the view known as *descriptivism*, which holds that the semantic value of a proper name is identical to that of some definite description associated with the name. It also contrasts with Carnap's view that the semantic value of a name is an *intension*, that is, a function from possible worlds to objects. So concerning semantics, Føllesdal and Kripke appear to be largely in agreement. In particular, both are fully committed to the rigidity of names and the principle of compositionality.

Meta-semantics, on the other hand, is concerned with what is involved in an expression's having the various semantic properties that it happens to have, such as its semantic structure and its semantic value. The expressions in question are purely syntactical items—ink marks on paper and vibrations in the air—and thus have no intrinsic semantic significance. What is it, then, that endows these intrinsically “dead” syntactic items with semantic structure and semantic values? These questions belong to meta-semantics rather than semantics proper. The relation between semantics and meta-semantics can be compared with that between economics and what we may call *meta-economics*. Economics is concerned with the laws governing money; for instance, that an excessive supply of money leads to inflation. Meta-economics, on the other hand, is concerned with what is involved in various objects' having monetary value; for instance, what

makes it the case that a piece of printed paper can be worth €100. Since neither semantic nor monetary properties are intrinsic to the items in question, there must be some account of what the possession of such properties consists in. This account is likely to draw on both psychological facts about the agents who operate with the items in question and sociological facts about these agents' interaction.

Our primary interest is in meta-semantic questions concerning singular terms. When a singular term refers to an object, what is the nature of this relation? What does this relationship consist in? The most influential answers to these meta-semantic questions are due to Kripke and Gareth Evans. According to Kripke, a name refers to its bearer in virtue of the historical chains by which the name has been passed down from people directly acquainted with the referent to the contemporary users of the name.³ People's beliefs about a name and its reference thus play an absolutely minimal role in Kripke's account of how the name comes to have its reference. Evans, on the other hand, chooses to build on the idea of information being transmitted in causal chains.

Føllesdal repeatedly and forcefully distances himself from such historical and causal theories of reference. He even writes that "my view [on the tie between singular terms and their objects] is much closer to Frege's than to Kripke's" (Føllesdal 1986, p. 109; see also Føllesdal 1997, p. 362). At first this claim may seem puzzling. How can Føllesdal, who like Kripke has vehemently criticized the descriptivist view of reference typically associated with Frege,⁴ now suddenly declare that he wants a view that is closer to Frege's than to Kripke's? But in fact there is absolutely no contradiction here. The distinction between semantics and meta-semantics helps us to see why. What Føllesdal has been defending, as against Frege, is the *semantic* thesis that names are rigid designators. But the passage just quoted in no way goes back on this. Rather, what Føllesdal says in this passage is that he wants a *meta-semantic* account of the link between a name and its

³ As Føllesdal 2004 pp. xxix-xxx points out, this view is anticipated in Geach 1969.

⁴ This interpretation of Frege has been challenged; see for instance Evans 1982 and McDowell 1977. For the purposes of this article I need not take a stand on this exegetical issue.

referent which is closer to Frege's than to Kripke's—while holding on to the semantic thesis that names are rigid designators.

The following desiderata for a meta-semantic theory of reference can be extracted from Føllesdal's writings.

1. *Generality*. The theory of reference must be general enough to cover reference to all kinds of objects, including abstract objects, with which there can be no direct acquaintance or causal interaction.⁵
2. *Fregean sense*. The theory must allow for some Fregean notion of sense; that is, for some mode of presentation of the referent.⁶
3. *Reference failure*. The theory must have something plausible to say about cases of reference failure, where the subject does his part of the job but the world fails to provide a unique referent. (Frege attempted to account for this by allowing a full-blown sense even in the absence of a referent.)⁷
4. *Individuation*. The theory must bring out the role that individuation plays in the determination of reference.⁸
5. *Cognitive Constraint*. In “non-parasitic” cases of reference, where a person refers to an object all by himself and not merely exploits the referential capacities of other members of his community, this person must know which object he is referring to; that is, he must possess some discriminating conception of the referent.⁹

⁵ See e.g. Føllesdal 2004, p. xxxi.

⁶ See e.g. Føllesdal 1986, p. 111 and Føllesdal 1997, p. 364.

⁷ See e.g. Føllesdal 1986, pp. 110-111; Føllesdal 1997, p. 359; and Føllesdal 2004, p. xxxi.

⁸ See e.g. Føllesdal 2004, p. xxxii.

⁹ See e.g. Føllesdal 1986, p. 109 and Føllesdal 2004, p. xxxi. The desideratum of Cognitive Constraint is related to what Evans 1982 calls *Russell's Principle*. But Cognitive Constraint is weaker than Russell's Principle because of its explicit restriction to non-parasitic cases of reference.

Summing up this section, I first subsumed Føllesdal's distinction between a semantics and a theory of reference under the more general distinction between semantics and meta-semantics. Then I explained how Føllesdal and Kripke agree on *the semantics* of singular terms: both defend the thesis that singular terms are rigid designators. Finally I explained how Føllesdal sharply disagrees with Kripke on *the meta-semantics* of singular terms: where Kripke defends his well-known historical account, Føllesdal wants a more Fregean theory of reference that satisfies the above five desiderata.

2. Towards a Fregean Theory of Reference

Frege's first serious investigation of the problem of reference appears to have been largely motivated by the desideratum of Generality. For at *Grundlagen* (Frege 1884) §62 Frege raises the following question. "How, then, are the numbers to be given to us, if we cannot have any ideas or intuitions of them?" The problem—which is still very much with us—is of course that numbers cannot be perceived or in any way experimentally detected. How then can we refer to such objects?

I claim that Frege's question at §62 belongs to what I have called meta-semantics. To see this, recall that by this stage of *Grundlagen* Frege has already argued that the numerals refer to objects (as opposed to second-order concepts) and rejected the psychologistic view that these objects are purely mental items. So Frege is at this stage entitled to assume his own platonistic view that the numbers are independently existing objects. Given this assumption, the semantic question what objects different numerals refer to has a completely straightforward answer; for instance, the numerals '7' and 'VII' refer to the number 7. Rather, Frege's present concern is with the meta-semantic question what facts about reference to natural numbers *consist in*. How can our numerals "latch on to" the natural numbers, given that there is no perception or causal interaction that can serve as a link?

Frege's next sentence proposes a way of addressing this hard meta-semantic question. "Since it is only in the context of a sentence that words have any meaning, our problem becomes this: To define the sense of a sentence in which a number word occurs." The doctrine that words have

meaning only in the context of a sentence has become known as *the Context Principle*.¹⁰ What Frege proposes is that the Context Principle has an essential role to play in the explanation of reference, both in general and to numbers and other abstract objects in particular. The idea is to translate the problem of explaining what it is for a singular term to refer into the problem of explaining what it is for certain complete sentences involving this term to be meaningful. I will now outline a Frege-inspired theory of reference which is based on this proposal. Although I believe Frege anticipated many aspects of this theory, he would probably have disagreed with other parts of it. But my present goal is systematic, not exegetical.

I begin by narrowing down the problem in two different ways. My first restriction is to focus on thought rather than on language.¹¹ For the purposes of this paper I will thus not attempt to say anything about how *linguistic expressions* come to refer but rather focus on the corresponding problem concerning *mental representations*.

My *explanandum* will thus be what is involved in someone's capacity for singular reference to various sorts of objects. The Frege-inspired proposal that I will investigate is that an adequate explanation of this capacity for singular reference will take the form of an explanation of what is involved in the person's capacity for understanding *complete thoughts* concerning objects of the sort in question.¹²

¹⁰ See also *ibid.* pp. x, 71, and 116. I have changed the translation of 'Satz' from 'proposition' to 'sentence'. This is reasonable, given that Frege talks about *words* occurring in a "Satz."

¹¹ In doing so I am to a large extent leaving the historical Frege behind, given his emphasis on starting with language rather than thought. But this departure is less radical than it may seem. In particular, my approach to thought (roughly that of Evans 1982) is very different from "the psychologistic" one that Frege so forcefully criticized: I seek to show how thought can have an objective propositional content, which is (at least in principle) intersubjectively accessible.

¹² Strictly speaking, I here collapse two steps. The first step is Frege's suggestion that questions concerning singular reference be addressed in terms of analogous questions concerning *complete thoughts*. In particular, in virtue of what does a physical state of an agent have a particular thought as its content? The second step is to approach this question about thoughts in terms of *the notion of understanding*. Doing so is quite nat-

This first restriction allows us to concentrate on an individual person rather than on a whole language community. This is a huge simplification. For instance, we can now hold that reference involves some Fregean mode of presentation but allow this mode of presentation to vary with each individual act of reference.¹³ In contrast, if a notion of sense is to be attached to an expression of a public language, then this sense will have to be shared by every competent speaker of this language.

My second restriction is to focus on *canonical* cases of singular reference.¹⁴ These are certain maximally direct ways of referring to objects, where the referent is “directly present” to the thinker. For instance, referring to a person whom I see immediately in front of me is canonical, whereas referring to Napoleon, with whom I am in no way acquainted, is not. (More examples of canonical reference will be presented shortly.) Having made these two restrictions, Frege’s proposal becomes the following: We can explain what is involved in someone’s capacity for canonical singular reference to objects of a certain kind by explaining what is involved in his or her capacity for understanding complete thoughts concerning such objects.

A slight simplification of this is possible. We observe that it makes sense to begin by explaining someone’s understanding of identity statements before attempting to explain his understanding of thoughts more generally. This strategy is adopted by Frege himself in *Grundlagen*.¹⁵ The rationale is

ural; for in order to stand in some propositional attitude to a thought, one presumably needs to understand that thought.

¹³ The resulting notion of mode of presentation may thus have more in common with Husserl’s notion of *noema* than with Frege’s notion of *Sinn*.

¹⁴ In the terminology of Evans 1982, my goal is to explain what our understanding of the relevant kind of “fundamental Ideas” consists in. Following Michael Dummett and Gareth Evans I believe non-canonical reference must be explained in terms of someone’s ability to recognize the referent when presented with it in a canonical way. See Dummett 1981, pp. 231-239 and Evans 1982, pp. 109-112.

¹⁵ Frege unfortunately abandons this strategy in *Grundgesetze*. For an analysis, see my Linnebo 2004.

that, before one can understand what it means for an object to possess properties and stand in relations, one needs to know how to distinguish the object from other objects and how to re-identify it when presented with it in alternative ways.¹⁶ When this observation is added, we arrive at what will be our official statement of Frege's proposal: We can translate the problem of explaining our capacity for canonical singular reference into the related but different problem of explaining our capacity for understanding identity statements concerning the object in question.

How should this proposal be carried out? Again Frege makes an ingenious suggestion. The core idea is that canonical reference has a rich and systematic structure. Firstly, objects are always presented to us only via some of their parts or aspects. And secondly, we have a grasp of how two such parts or aspects must be related for them to pick out the same object. Here are some examples.¹⁷

1. *Physical bodies*.¹⁸ A physical body is most directly presented in perception, where we causally interact with one or more of its spatio-temporal parts. Two such parts determine the same physical body just in case they are connected through a continuous stretch of solid stuff, all of which belongs to a common unit of motion.¹⁹
2. *Directions*. A direction is most directly presented by means of a line (or some other directed object) that has the direction in question. Two lines determine the same direction just in case they are parallel.

¹⁶ Cf. Evans 1982, who explains fundamental Ideas in terms of "fundamental grounds of difference."

¹⁷ In a more complete treatment, each example would of course have to be developed in greater detail and defended against objections. My present goal is merely to sketch some promising examples in order to illustrate how the Fregean framework functions.

¹⁸ By "physical body" I mean, roughly, a cohesive physical object with natural boundaries. Apples and oranges are thus paradigmatic physical bodies. By contrast, proper parts of an apple and arbitrary mereological sums of apple stuff may be physical *objects* but fail to qualify as physical *bodies*.

¹⁹ I elaborate on this view and defend it against some natural objections in my Linnebo 2005.

3. *Shapes*. This case is analogous to that of directions: Shapes are most directly presented by things or figures that have the shape in question. Two such things or figures determine the same shape just in case they are congruent.
4. *Syntactic types*. Syntactic types are most directly presented by means of their tokens. Two tokens determine the same type just in case they count as equivalent (“equi-typical”) according to the relevant standards.
5. *Natural numbers*. A natural number is most directly presented by means of some member of a sequence of numerals. Two numerals determine the same number just in case they occupy analogous positions in their respective sequences.²⁰

These examples suggest that canonical cases of singular reference are always based on two elements. First, there is an intermediary entity in terms of which the referent is most immediately presented. Let’s call this the *presentation*. Second, there is a relation which specifies the condition under which two presentations determine the same referent. Let’s refer to this as the *unity relation*. Finally, let’s call an ordered pair $\langle u, \approx \rangle$ consisting of a presentation u and a unity relation \approx applicable to this presentation a *referential attempt*. Frege’s proposal is then that canonical reference is based on referential attempts. Does this proposal yield an adequate explanation of what someone’s understanding of singular reference consists in?

A *formal* adequacy condition is obviously that the account be non-circular. It is easily seen that the form of our proposal allows it to be non-circular.

²⁰ Related ideas are found in Parsons 1971. This view of the natural numbers as finite *ordinals* contrasts with the logicist view that the natural numbers are finite *cardinals*, individuated by Hume’s Principle (which says that two numbers are identical just in case the concepts whose numbers they are are equinumerous). However, both views are compatible with the Fregean account of reference. It is thus largely an empirical question which view best describes human thought about the natural numbers.

Consider for instance the case of directions. What I have proposed is that someone's understanding of an identity statement concerning directions can be explained in terms of his being suitably related to lines (in terms of which directions are presented) and having a suitable grasp of parallelism (which is the unity relation). In this case there is no threat of circularity; for we can explain what it is for someone to be suitably related to lines and to have a suitable grasp of parallelism without presupposing any prior ability to explain reference to directions. Next, we observe that there is nothing in this example that is peculiar to the case of directions. My proposal is to explain someone's understanding of an identity statement in terms of this person's being suitably related to the relevant presentations and having a suitable grasp of the relevant unity relation. This explanation will of course have to include an account of what it is for a person to be suitably related to these presentations and to have a suitable grasp of this unity relation. But there is no general reason why this account should presuppose what we are trying to explain, namely reference to the sort of objects that are determined by these presentations and this unity relation.²¹

The *material* adequacy condition is that the account should capture what someone's capacity for singular reference *consists in*. My argument that this adequacy condition is satisfied is based on two claims: first, that my account explains what the subject's understanding of identity statements involving the referent consists in; and second, that this understanding explains the subject's capacity for singular reference to the object in question.

Let's begin with the first claim. Consider a representation *a* purporting to make singular reference to some object. According to my account, this representation is associated with some referential attempt $\langle u, \approx \rangle$, which specifies how the referent is presented and when two such presentations deter-

²¹ This is of course not to say that there cannot be *particular* cases where such an illicit presupposition exists. In fact, elsewhere I suggest that some of the problems encountered by Frege's proposal are caused by the use of presentations and unity relations an adequate grasp of which *would* presuppose an ability to refer to the entities in question, thus making the account viciously circular. See Linnebo 2006, pp. 166-168 and Linnebo forthcoming, Section 2.2.

mine the same referent. By operating with this referential attempt, the subject will be able to understand any thought of the form $\lceil \mathbf{a} = \mathbf{b} \rceil$, where \mathbf{b} is any other representation purporting to make singular reference to an object of the kind in question. For according to my account, \mathbf{b} too must be associated with some referential attempt, say $\langle v, \approx \rangle$. Moreover, we are assuming that the subject operates correctly with these representations, namely in accordance with the following principle for the identity of their semantic values:

$$(SV) \quad [\mathbf{a}] = [\mathbf{b}] \leftrightarrow u \approx v$$

This means that the subject has an ability to track the referent of \mathbf{a} and to distinguish it from other objects of the same sort.

My second claim is that this competence is naturally described as knowing (in a non-parasitic way) what object the representation \mathbf{a} refers to. Consider for instance the case of physical bodies. Assume someone is digging in the garden, hits upon something hard with her shovel, and as a result forms the thought: This body is large. Later she hits upon something hard again, one meter away from the first encounter, and as a result forms the thought: This body is identical to that body. Finally, our subject appreciates that this identity statement is true just in case the two chunks of solid stuff that she has hit upon are spatiotemporally connected in the suitable way. It is extremely plausible to describe this capacity as a capacity to refer to physical bodies. For instance, if a robot was equipped with perception-like mechanisms and programmed so as to operate with the appropriate unity relation, it would make sense to ascribe to the robot a basic capacity for referring to physical bodies.

On the view that I am advocating, the unity relation \approx implicitly defines a (partial) function f_{\approx} that maps a presentation u to the referent, if any, that u picks out. This is encapsulated in what I will call *principles of individuation*:

$$(PI) \quad f_{\approx}(u) = f_{\approx}(v) \leftrightarrow u \approx v$$

Of course, when formulating principles of individuation, we philosophers make use of our own ability to refer to objects of the kind in question. But this is perfectly permissible. We are allowed to presuppose *that* we can refer to objects of the kind in question. What we are not allowed to presuppose is an explanation of what this ability *consists in*. But no such presupposition is made.

3. Assessing the Fregean Theory of Reference

Recall from Section 1 the kind of account that Føllesdal wants of the phenomenon of reference. Firstly, he wants a *semantics* that incorporates both the principle of compositionality and his own discovery that names are rigid designators. Secondly, he wants a *meta-semantic account* (or what he calls a “theory of reference”) that satisfies the five desiderata that I identified. I will now assess the extent to which the Frege-inspired theory of reference outlined in the previous section delivers what Føllesdal wants.

Obviously, my theory can at best be the beginning of the sort of theory Føllesdal wants. For I have explicitly restricted my attention in two ways: firstly to reference at the level of thought rather than at the level of language; and secondly, to canonical cases of such reference. By contrast, Føllesdal is concerned with the reference of linguistic expressions, including cases that can in no way be regarded as canonical. My question can thus only be whether my theory, *as far as it goes*, has those features that Føllesdal wants.²² If it does, then this theory may serve as a stepping stone for later attempts to develop the more ambitious sort of theory that Føllesdal ultimately wants.²³

²² Henceforth, this qualification will tacitly be assumed to be in place unless otherwise stated.

²³ Should it, on the other hand, turn out that no such ambitious theory is possible, then this would not automatically threaten my less ambitious theory. For core cases of some phenomenon may well enjoy a particularly nice explanation although this explanation cannot be extended to more peripheral cases.

3.1. The meta-semantic desiderata

It is fairly straightforward to see that my Frege-inspired theory of reference satisfies the five meta-semantic desiderata listed at the end of Section 1. The first desideratum—Generality—is clearly satisfied. Indeed, the theory was explicitly designed so as to be able to accommodate reference to all kinds of objects, including abstract ones.

Next, we observe that the theory incorporates a notion of *Fregean sense* or mode of presentation of the referent. For according to this theory, canonical reference to an object is mediated by a presentation and a unity relation. Note that this notion of sense is not spelled out in a descriptivist manner; that is, it does not identify the semantic contribution of a proper name with that of some description.²⁴ Rather, on my account the mode of presentation is a feature of the meta-semantic mechanism by which a semantically simple item comes to refer to an object. A mode of presentation is thus a part of *an explanation* of how an item comes to possess a particular semantic content, but not part of *this content itself*.

Further, I claim that the theory provides a plausible analysis of what is going on in cases of *reference failure*. For according to my theory, there will be presentations that fail to determine referents. There are for instance spatiotemporal parts that fail to determine unique bodies. For instance, if I point to the floor and say ‘this body’, I will probably fail to determine a unique body.²⁵

²⁴ For other attempts to articulate a non-descriptivist notion of sense, see Evans 1982 and McDowell 1977.

²⁵ The possibility of reference failure has consequences for the unity relation. Since a unity relation gives the condition under which two presentations determine the same referent, this relation will always have to be symmetric and transitive. However, a unity relation will only be reflexive on those presentations that succeed in determining referents. For a referential attempt $\langle u, \approx \rangle$ succeeds in determining a referent just in case the presentation u bears the unity relation \approx to itself. From this it also follows that the function f_{\approx} that figures in the principles of individuation is a partial function whose domain is identical to the field of the unity relation \approx . (Recall that the *field* of a relation R is the set of objects which R relates. Thus, when R is dyadic, its field is the set $\{x \mid$

The final two desiderata were Individuation—which requires that the notion of individuation play a role in the determination of reference—and Cognitive Constraint—which says that for a thinker to refer non-parasitically to an object, he must know which object he is referring to. These desiderata are clearly satisfied as well.

3.2. The principle of compositionality

I now turn to a worry about the compatibility of my Fregean theory of reference with the principle of compositionality. This worry is based on the following observation. According to the principle of compositionality, the meaning of a complex expression is to be explained in terms of the meanings of its constituent parts. So here we are explaining a property of a complex expression in terms of the properties of its simple constituents. But according to my Fregean theory of reference, the referentiality of a singular representation is partially explained in terms of the meaningfulness of identities involving this representation. So here we are explaining a property of a semantically simple item in terms of the properties of more complex expressions of which this simple is a constituent. This means that “the order of explanation” associated with my Fregean theory of reference is the opposite of that dictated by the principle of compositionality. It thus appears that my theory conflicts with the important principle of compositionality.²⁶

I would like to suggest that our distinction between semantics and meta-semantics provides the key to resolving the apparent conflict. The principle of compositionality is concerned with the assignment of semantic values to complex expressions and thus belongs to semantics. My Fregean theory of reference, on the other hand, is concerned with what is involved in an ex-

$\exists y (Rxy \vee Ryx)$. Note that it is a theorem of first-order logic that a symmetric and transitive relation is reflexive on all objects in its field.)

²⁶ One way *not* to resolve this apparent conflict would be by claiming that, whereas my theory is only concerned with thought, the principle of compositionality governs only linguistic meaning. This response is unacceptable because an analogous principle of compositionality is supposed to apply to the contents of thoughts.

pression's having the various semantic properties it happens to have and thus belongs to meta-semantics. Since the principle of compositionality and our Fregean account of reference have completely different concerns, there is no conflict. What the principle of compositionality says is that the semantic value of a complex expression is determined by the semantic values of its simple constituents. But it says nothing about how other kinds of explanation—such as meta-semantic explanations of what it is for an expression to have a semantic value in the first place—should proceed.

However, the most popular response among philosophers who seek to use some form of Frege's Context Principle to explain reference has been to concede that the apparent conflict is genuine and therefore to argue that the principle of compositionality has to be rejected or at least weakened.²⁷ But rejecting or weakening the principle of compositionality is obviously a steep price to pay. Why, then, have so many philosophers found this response inevitable?

I believe the answer has to do with a dangerous ambiguity in the wording of Frege's proposal. "Since it is only in the context of a sentence that words have any meaning," Frege writes, "our problem becomes this: *To define the sense of a sentence in which a number word occurs.*"²⁸ This is ambiguous between a semantic and a meta-semantic reading. On the semantic reading, our task becomes to *specify* the meaning or sense of identity statements in which number words occur. But on the meta-semantic reading, our task is to *explain what makes it the case that such identity statements have the meanings that they happen to have.*

Frege's proposal has traditionally been interpreted along the lines of the semantic reading. It is then natural to assume that what Frege proposes is that the meaning of problematic identity statements be given by a "reductive" truth-condition

(T-Red) $\ulcorner \mathbf{a} = \mathbf{b} \urcorner$ is true iff $u \approx v$

²⁷ See Hale 1997 and Wright 1997.

²⁸ *Grundlagen* §62; my italics.

where **a** and **b** are representations associated with referential attempts $\langle u, \approx \rangle$ and $\langle v, \approx \rangle$ respectively. On this reading there will indeed be a conflict with the principle of compositionality. For according to this principle, the semantic value of an atomic sentence $\mathbf{P}(\mathbf{a}_1, \dots, \mathbf{a}_n)$ is functionally determined as $[\mathbf{P}]([\mathbf{a}_1], \dots, [\mathbf{a}_n])$. Applied to the identity $\lceil \mathbf{a} = \mathbf{b} \rceil$, this yields a different, completely trivial truth-condition:

(T-Triv) $\lceil \mathbf{a} = \mathbf{b} \rceil$ is true iff $[\mathbf{a}] = [\mathbf{b}]$

(And any further or alternative semantic analysis is out of the question, given that the terms **a** and **b** are supposed to be semantically simple.) Moreover, the truth-condition (T-Triv) will be of absolutely no use in the project of explaining some problematic form of reference. For the right-hand side of (T-Triv)—unlike that of (T-Red)—involves precisely the sort of reference that we are attempting to explain.

Faced with this choice between the reductive truth-condition (T-Red), which allows the explanatory project to progress, and the trivial one (T-Triv), on which the explanation cannot even get started, it is of course tempting to insist that it is the former that gives the meaning of the identity statement, and that if this conflicts with the compositionality of meaning, then so much the worse for this principle of compositionality. This appears to have been Frege's view in *Grundlagen*, where he talks about the right-hand side being a "recarving" of the meaning of the left-hand side.²⁹ This "recarving thesis" is explicitly endorsed by prominent contemporary defenders of Fregean ideas about reference, such as Bob Hale and Crispin Wright.³⁰

However, I have insisted throughout this paper that Frege's proposal is of meta-semantic nature. I am therefore under no pressure to say that (T-Red) gives *the meaning* of an identity statement. I can instead maintain that the

²⁹ See Frege 1884, §64.

³⁰ See Hale 1997 and Wright 1997.

only semantically generated truth-condition for an identity statement is the trivial one (T-Triv). What Frege proposes is rather an account of what a subject's *understanding of an identity statement consists in*. And as we have seen, this account involves the principle for the identity of their semantic values:

$$(SV) \quad [\mathbf{a}] = [\mathbf{b}] \leftrightarrow u \approx v$$

When this principle (SV) is combined with the trivial truth-condition (T-Triv), we do indeed get the reductive one (T-Red), which now emerges as a hybrid of semantic and meta-semantic facts.

3.3. The rigidity thesis

A final worry concerns the compatibility of my Fregean theory of reference with the semantic thesis that names and their mental counterparts are rigid designators. Consider a representation \mathbf{a} associated with a referential attempt $\langle u, \approx \rangle$. Let f_{\approx} be the function determined from \approx in accordance with the Principle of Individuation (PI). I have argued that \mathbf{a} refers, if at all, to the object $f_{\approx}(u)$. One may then wonder whether my view doesn't collapse back into some version of the descriptivist view of names criticized by Kripke and Føllesdal. Specifically, am I not committed to identifying the meaning of \mathbf{a} with that of the description "the f_{\approx} of u ," with the result that \mathbf{a} isn't a rigid designator after all? (For instance, one chunk of physical matter may be part of different bodies in different possible worlds.)

The above discussion provides the resources needed to respond to this worry. On my proposal, the nature of the function-argument structure $f_{\approx}(u)$ is entirely meta-semantic, not semantic. The expression \mathbf{a} is semantically simple, and its semantic value, if any, is just *the object* $f_{\approx}(u)$. How this referent is determined is a meta-semantic matter, of no immediate semantic significance. As far as semantics is concerned, \mathbf{a} is a simple term or representation whose semantic value is just an object. More generally, not every kind of structure involved in the phenomenon of reference is *semantic* structure. For instance, reference is often based on perception, and perception is undoubtedly a complicated process that involves all kinds of struc-

turing of sensory information. But this structure will generally not be semantic structure. Although perception is often *presupposed by* the relation of reference and thus also by semantics, perception and its structure aren't thereby *included in* semantics.

My claim that the function-argument structure $f_{\approx}(u)$ isn't semantic structure enjoys independent evidence as well. Semantic structure is by and large accessible to consciousness; otherwise we wouldn't know or be rationally responsible for what we say and think. But someone can understand reference to shapes and bodies without having any conscious knowledge of how such reference is structured. Someone's competence with this structure may be located entirely at a "subpersonal" level, much as the structuring involved in perception is. This is evidence that the function-argument structure $f_{\approx}(u)$ isn't semantic. And if that is right, then my account will be fully compatible with the rigidity thesis and in no danger of collapsing back into descriptivism.

My conclusion is thus that my Fregean theory of reference manages to combine the semantic theses of rigidity and compositionality with Føllesdal's meta-semantic desiderata. Although my theory is much more limited in scope than what Føllesdal ultimately wants, this is at least a first step in that direction. Moreover, this establishes that there is no inherent conflict between a semantic theory that takes singular terms (or their mental counterparts) to be rigid designators and a meta-semantic theory that satisfies Føllesdal's five Frege-inspired desiderata. Although Føllesdal (like Kripke) takes himself to break radically with Frege on the *semantics* of singular terms, there is room for extensive agreement with Frege (and disagreement with Kripke) on the *meta-semantics* of singular terms.³¹

³¹ Thanks to Anthony Everett for comments on an earlier version of this paper and to the participants at the Lauener Prize Symposium for discussion. I am particularly grateful to Føllesdal for extensive discussions of the ideas of this paper, both at the symposium and after.

References

- Dummett, Michael. 1981. *Frege: Philosophy of Language*, 2nd ed. (Cambridge, MA: Harvard University Press)
- Evans, Gareth. 1982. *Varieties of Reference* (Oxford: Oxford University Press)
- Føllesdal, Dagfinn. 1986. "Essentialism and Reference," in L.E. Hahn and P.A. Schilpp (eds.), *The Philosophy of W.V. Quine* (La Salle, IL: Open Court), pp. 97-113
- . 1997. "Conceptual Change and Reference," in C. Hubig (ed.), *Cognitio Humana: Dynamik des Wissens und der Werte* (Berlin: Akademie Verlag), pp. 351-367
- . 2004. *Referential Opacity and Modal Logic* (New York and London: Routledge)
- Frege, Gottlob. 1884. *Foundations of Arithmetic*. Transl. J.L. Austin (Oxford: Blackwell, 1953)
- Geach, Peter. 1969. "The Perils of Pauline," *Review of Metaphysics* 23, pp. 287-300
- Hale, Bob. 1997. "Grundlagen §64," *Proceedings of the Aristotelian Society* 97(3), pp. 243-261; repr. with a postscript in Hale and Wright 2001
- and Crispin Wright. 2001. *Reason's Proper Study* (Oxford: Clarendon)
- Linnebo, Øystein. 2004. "Frege's Proof of Referentiality," *Notre Dame Journal of Formal Logic* 45(2), pp. 73-98
- . 2005. "To Be Is to be an *F*," *Dialectica* 59(2), pp. 201-222
- . 2006. "Sets, Properties, and Unrestricted Quantification," in A. Rayo and G. Uzquiano (eds.), *Absolute Generality* (Oxford: Clarendon)
- . 2009. "Frege's Context Principle and Reference to the Natural Numbers," in S. Lindström et al. (eds.), *Logicism, Intuitionism, and Formalism: What Has Become of Them?* (Springer), pp. 47-68

- McDowell, John. 1977. "On the Sense and Reference of a Proper Name," *Mind* 86(342), pp. 159-195
- Parsons, Charles. 1971. "Ontology and Mathematics," *Philosophical Review* 80, pp. 151-176; repr. in his *Mathematics in Philosophy* (Ithaca, NY: Cornell University Press)
- Stalnaker, Robert. 1997. "Reference and Necessity," in B. Hale and C. Wright (eds.), *Blackwell Companion to the Philosophy of Language* (Oxford: Blackwell), pp. 534-554; repr. in his *Ways of World Might Be* (Oxford: Clarendon, 2003)
- Wright, Crispin. 1997. "The Philosophical Significance of Frege's Theorem," in R. Heck (ed.), *Language, Thought, and Logic: Essays in Honour of Michael Dummett* (Oxford: Clarendon); repr. in Hale and Wright 2001

