Compositionality and Frege's Context Principle

Øystein Linnebo

University of Bristol

Draft of 24 January 2008; slightly edited September 4, 2008

Abstract

Two important principles in the philosophy of language are due to Frege: the context principle, which says that words have meaning only in the context of a sentence, and the principle of compositionality, which says that the meaning of a sentence is determined by the meanings of its constituents. Many philosophers have taken there to be a conflict between these two principles. I challenge this widespread view and show that the two principles are fully compatible. My reconciliation of the two principles is based on a distinction between two kinds of explanation of meaning: semantic and meta-semantic. I also show how this opens the way for an informative account of reference which is general enough to subsume even reference to mathematical objects.

1 Introduction

It is widely agreed that one of the most important principles of Frege's philosophy is his *context principle*, which urges us never to "ask for the meaning of a word in isolation, but only in the context of a sentence" ((Frege, 1953), p. x). But this is about as far as the agreement goes. There has been a great deal of controversy about what exactly the context principle says and whether it is defensible. In this article I develop and defend an interpretation of the context principle according to which its main function is to urge that the relation of reference be explained not in isolation but rather in the context of complete sentences. I also argue that the context principle plays an essential role in Frege's logicist account of mathematics, where it is meant to explain how it is possible to refer to the natural numbers and thus entertain arithmetical thoughts.

This interpretation raises the question whether the context principle is compatible with another important principle in the philosophy of language with an equally noble Fregean pedigree, namely *the principle of compositionality*. According to this latter principle, the meaning of a complete sentence must be explained in terms of the meanings of its subsentential parts, including those of its singular terms. The two Fregean principles thus call for explanations that proceed in opposite directions: the context principle calls for an explanation of the subsentential in terms of the sentential; and the principle of compositionality, for an explanation of the sentential in terms of the subsentential. Are these two principles compatible? Many philosophers—including the Frege of *Grundlagen* and prominent followers such as Michael Dummett and Bob Hale—have thought that there is at least a partial conflict between the two principles.¹

A major goal of this paper is to challenge this wide-spread view and show that the two principles are fully compatible after all. If correct, this will be a significant result. It will obviously solve the exegetical puzzle of how Frege can have been committed to both principles. But more importantly, it will show how two *prima facie* attractive principles can be reconciled. The context principle is attractive because it opens the way for an informative account of reference which is general enough to subsume even to reference to mathematical objects. And the principle of compositionality is attractive because it is an integral part of many of our best semantic theories.

My reconciliation of the two principles is based on a careful distinction between two different kinds of explanation of meaning. The context principle urges us to explain the reference of a singular term by explaining the meanings of complete sentences in which the term occurs. But what *kind of explanation* of sentential meaning is required? According to the received view, which derives from Frege's *Grundlagen*, what is required is a *restatement* of the meanings of the sentences in question which avoids all use of the singular terms whose reference is to be explained. These restatements are said to be "recarvings" of the meanings of the sentences in question: they have the same meanings as the original sentences but avoid all use of the problematic terms. In particular, the right-hand sides of so-called "abstraction principles" are said to "recarve" the meanings of their left-hand sides. I argue that this sort of recarving of meaning conflicts with a plausible version of the principle of compositionality.

However, I reject the idea that any sort of recarving of meanings is needed in the first place. I argue that this idea is based on a misunderstanding of the kind of explanation of sentential meaning that the context principle requires. What the principle requires is not a *restatement* of the meanings of sentences containing problematic terms in different and less problematic terms but an *account of what it is for these sentences to have the meanings that*

¹See e.g. (Dummett, 1991b), esp. ch.s 14-16; (Frege, 1953); (Hale, 1994); and (Hale, 1997).

they have. We need to explain what having these meanings consists in. And this question can be answered without making any problematic claims about the recarving of meanings. The context principle can thus be purged of all talk about the recarving of meanings, and with it, of everything that conflicts with the principle of compositionality. This results in an attractive version of the context principle, which holds great promise for the explanation of reference, both in general and to natural numbers and other abstract objects in particular.

The paper is organized as follows. First I explain the context principle in light of the role it is supposed to play in Frege's *Grundlagen* (Section 2). Then I explain how a particular interpretation of the context principle has led Frege and many others to endorse a strong thesis about the recarving of meanings (Section 3). Next I show how this thesis about recarving conflicts with a strong form of the principle of compositionality (Section 4) as well as with the strategy set out in *Grundlagen* for explaining reference by means of the context principle (Section 5). However, when Frege in the 1890s began to defend the principle of compositionality, he gave up on the idea of recarving meanings, while still holding on to the context principle. I articulate an alternative interpretation of the context principle which shows how this is possible (Section 6). I end with some remarks about the prospects for an account of reference based on this alternative interpretation of the context principle (Section 7).

2 The role of the context principle in *Grundlagen*

Frege's context principle has received a bewildering range of interpretations.² But in order to understand the principle, there is no better approach than to examine the role it plays in the work of Frege's where it figures most prominently, namely his *Grundlagen*. Since the main goal of *Grundlagen* is to defend a logicist account of mathematics, I will be particularly concerned with the contribution that the context principle is supposed to make to this account of mathematics.

The most famous occurrence of the context principle is probably in the following passage.³

How, then, are the numbers to be given to us, if we cannot have any ideas or intuitions of them? Since it is only in the context of a sentence that words have

²See (Pelletier, 2001) for an overview.

³(Dummett, 1978), p. 38 characterizes this as "probably the most important philosophical statement Frege ever made." Later he uses even stronger words, describing it as "arguably the most pregnant philosophical paragraph ever written" ((Dummett, 1991b), p. 111). Although this is a bold claim, it is not easy to think of a better example.

any meaning, our problem becomes this: To explain the sense of a sentence in which a number word occurs. ((Frege, 1953), $\S62$)⁴

Three things happen in this short passage. First a problem is described. Then the context principle is stated. Finally Frege proposes that the context principle opens for a solution to the problem. Let's consider these three things in order.

The problem is introduced in the opening sentence. How is reference to natural numbers possible? This is today a familiar question, in large part thanks to (Benacerraf, 1973). The natural numbers are supposed to be abstract and thus incapable of participating in causal processes. But this rules out any kind of perceptual access to numbers, since perception is always based on causal processes. For the same reason it rules out any kind of experimental detection of numbers. How can we then refer to natural numbers and gain knowledge of them?

Some philosophers move very quickly from these kinds of consideration to the claim that reference to numbers and other kinds of abstract objects is impossible. But this inference is unwarranted. The right lesson to draw is rather that one of our best understood models of how reference comes about—that based on immediate perceptual access—is inapplicable, as one of its key ingredients is missing. But for all we know, there may be other models which are better suited for explaining reference to abstract objects. In fact, even in cases of reference to concrete objects, it is highly doubtful that a causal connection can provide a complete explanation. For how can a causal connection single out the intended referent, as opposed to its surface, one of its temporal parts, or the mereological sum of its atoms?⁵ I will later suggest that reference to concrete objects relies not just on a causal connection but also on *criteria of identity*. This will open for the view that reference to abstract objects is just a limiting case, where criteria of identity are still present but the causal component has disappeared completely.

The next sentence introduces the context principle itself: "it is only in the context of a sentence that words have any meaning." This appears to make the extremely strong claim that that words never have any meaning in isolation and that sentential contexts have a complete of monopoly on meaning:

 $^{^{4}}$ My translation differs slightly from that of J.L. Austin in that I render Frege's original 'Satz' and 'erklären' as respectively 'sentence' (rather than 'proposition') and 'explain' (rather than 'define'). I do not regard this as controversial. In particular, the context in which a syntactic item such as a word occurs is clearly that of a sentence, not a proposition.

⁵This last example assumes that the referent is something different from the mereological sum of its atoms.

Sentential Monopoly: Whenever a word has meaning, it occurs in the context of a meaningful sentence.

So according to this thesis a word has no meaning whatsoever when it occurs outside the context of a sentence.

The representational powers of the parts of a painting provide a useful analogy. A black speck of paint in David's *The Coronation of Napoleon* may for instance represent Napoleon's left eye. But the representational power of the speck of paint seems to arise only in the context of the painting and not pertain to the speck itself in a way that is independent of this context. If the black speck has any context-independent representational power at all, this will be very limited—perhaps just to represent black things. According to Sentential Monopoly,⁶ the meanings of words are much like the representational powers of specks of paint: words do not have any meanings by themselves but become meaningful only in the context of complete sentences.

However, the analogy between words and specks of paint which is suggested by Sentential Monopoly is not a good one. Words typically do have fairly strong context-independent representational powers. For instance, the word 'eye' tends to represent eyes throughout contexts in which English is spoken. In fact, Sentential Monopoly contradicts one of the most important and least controversial principles of linguistics, namely that words have lexical meanings which play a role in determining the meanings of sentences in which they occur. This is also a principle to which Frege himself was firmly committed throughout most of his career. Frege's mature work provides some clear examples, some of which will be discussed in Section 6.1 below. But already in Frege's early work we find passages that hint at this principle. For instance, in the unpublished article "Boole's logical calculus and the Concept-scrip" from 1880-81 Frege writes that in his logically perfect language (*Begriffsschrift*) the "designations" of properties and relations

never occur on their own, but always in combinations which express contents of possible judgment. I could compare this with the behavior of the atom: we suppose an atom never to be found on its own, but only combined with others, moving out of one combination only in order to enter immediately into another.

 $((Frege, 1979), p. 17)^7$

⁶Capitalization will be used solely to refer to claims which have been explicitly defined in boldface.

⁷Frege can be excused for not knowing that Helium and other inert gases falsify his chemical supposition: for Helium was isolated only in 1895. In any event, this is irrelevant to Frege's logical point.

This passage suggests a view that is much weaker and more plausible than Sentential Monopoly. Words have independent representational powers, much like atoms have independent natures. But these representational powers typically cannot be manifested on their own but only in the context of sentences, much like atoms typically cannot occur on their own but only bound together in molecules.

The *Grundlagen* contains three other formulations of the context principle, none of which requires any commitment to Sentential Monopoly.

never ask for the meaning of a word in isolation, but only in the context of a sentence (p. x)

It is enough if the sentence taken as a whole has sense; it is this that confers on its parts also their content. $(\S 60)$

We next laid down the fundamental principle that we must never try to explain the meaning of a word in isolation, but only as it is used in the context of a sentence. $(\S106)$

The second of these formulations gives a very clear expression of a claim about the determination of meaning, which we can formulate as follows.

Sentential Determination: The meaning and logical form of a sentence determine the meanings of its constituent expressions.⁸

And the third (and to some extent the first) formulation of the context principle indicate instead a principle about explanatory priority.

Sentential Priority: The meanings of words cannot be explained in isolation

but must always be explained in the context of complete sentences.⁹

⁸It is not clear from the quoted passage whether the meaning of the sentence provides a sufficient basis for this determination or whether its logical form is needed as well. I have chosen the second interpretation because it is weaker and because (as we will see shortly) this suffices for the use to which the principle is put. (Dummett, 1956) interprets the context principle as making a claim related to Sentential Determination, only weaker and about sense rather than meaning, namely that "when I know the sense of all the sentences in which a word is used, then I know the sense of the word" (p. 39). But as Dummett notes, the antecedent of this conditional is much too strict and thus in need of refinement. See also (Dummett, 1981b), p. 405 for a version of Sentential Determination where meaning is understood as reference.

⁹This principle, with 'meaning' understood as *Fregean sense*, is Dummett's interpretation of the context principle in (Dummett, 1981a), chapter 1. But see also pp. 495-7 and (Dummett, 1995) for the ascription to Frege of the analogous principle where 'meaning' is understood as *reference*. See also (Dummett, 1981b), ch. 19.

So the official statements of the context principle in the *Grundlagen* leave it quite unclear what the principle actually says. In order to make progress, I propose that we look instead at the philosophical work that the context principle is supposed to do in *Grundlagen*.

This brings us to the third thing that happens in the passage quoted at the beginning of this section. Having introduced the context principle, Frege proposes that the principle has an essential role to play in the explanation of reference. Given the context principle, he claims, "our problem [of explaining how reference to numbers is possible] becomes this: To explain the sense of a sentence in which a number word occurs." Frege thus claims that the reference of a singular term must be explained in terms of the meanings of complete sentences in which the term occurs. This will translate a problem at the sub-sentential level—that of explaining the reference of a singular term—into a problem at the sentential level—that of explaining the meanings of complete sentences. His proposed use of the context principle can thus be stated as follows.

Frege's Proposal: The explanation of the reference of a singular term must proceed via an explanation of the meanings of complete sentences in which this term occurs.

Note that this proposal is just the result of applying Sentential Priority to singular terms and taking the meanings of singular terms to be their reference. It is thus fairly clear that it is Sentential Priority which is at the heart of Frege's project in *Grundlagen*. But as we will see shortly, Sentential Determination too plays a role in the implementation of this strategy. By contrast, the most problematic of the theses suggested by Frege's various formulations of the context principle—Sentential Monopoly—plays little or no role in the overall argument of the *Grundlagen*. I therefore doubt that Frege ever had any strong and deep commitment to Sentential Monopoly as formulated above.¹⁰ In fact, Sentential Monopoly will play no role whatsoever in the remainder of this paper.

3 The Recarving Thesis

Many questions remain concerning Frege's Proposal. Let me mention three.

The first question concerns the notion of meaning that is at play. In *Grundlagen* Frege never says much about the notion of meaning which he relies on but rather treats it in an

¹⁰Here I follow a long tradition which includes (Dummett, 1956); (Dummett, 1981a), ch. 1; and (Dummett, 1981b), ch. 19.

informal way. But in later works Frege of course had a lot to say about meaning. Starting in the 1890s he distinguishes sharply between sense and reference and articulates important insights about these notions and their interrelation. In this section I will follow *Grundlagen* and operate with one informal notion of meaning rather than with the two technical notions of sense and reference. But in later sections it will become necessary to consider more precise technical notions. Throughout the paper the word 'meaning' will be reserved for the informal notion, and the words 'sense' and 'reference' for Frege's two technical notions of *Sinn* and *Bedeutung*.

The second question concerns the class of sentences for which an explanation of meaning must be provided. Must we explain the meanings of *all* sentences involving "number words"? Or is it enough to provide such an explanation for some privileged subclass of such sentences? This is the question to which Frege most immediately turns his attention after the passage quoted at the beginning of the previous section. Recall that by Section 62 Frege has already settled that number words stand for "self-subsistent objects." This "gives us," he writes, "a class of sentences which must have a [meaning], namely those that express our recognition of a number as the same again."¹¹ Frege thus recommends that we focus on identity statements involving number words. This amounts to the recommendation of a major simplification of our task: let's begin by explaining the meanings of the appropriate identity statements. This simplification seems perfectly sensible. For it is hard to see how we can explain the meanings of more complicated sentences about numbers before we can explain something as simple and important as identity statements.¹²

We can in fact be even more specific about the class of sentences whose meanings are to be explained first. For as anyone familiar with the earlier parts of the *Grundlagen* will know, the "number words" that Frege has in mind have a very specific form. Frege has argued that number ascriptions are made first and foremost to *concepts*. For instance, the statement 'there are eight planets' should be analyzed as an ascription of the number eight to the concept *is a planet*. In accordance with this analysis Frege takes the basic number words to be ones of the form 'the number of Fs', where F is a concept term. The identity statements which Frege

¹¹In accordance with the terminological convention just adopted, I have substituted 'meaning' for 'sense' (or, in the original German, 'Sinn'), as this occurrence of the word relies on an information notion of meaning. ¹²Housever, in his formous "proof of referentiality" in *Computers* 1 (Free 1064) \$20.21. Free appears to

¹²However, in his famous "proof of referentiality" in *Grundgesetze* I (Frege, 1964) §§29-31, Frege appears to attach no special significance to identity statements but rather to hold that, in order to ensure that a singular term refers, *all* sentences in which the term occurs must be assigned truth-conditions. For a discussion of Frege's proof of referentiality, see my (Linnebo, 2004).

urges us to focus on are thus the ones flanked by number terms of this sort. Let's call these the *basic numerical identities*.

The third question concerns the nature of the explanation of meaning which Frege calls for. He proposes that the context principle will allow us to explain the reference of a "number word" by means of "explaining the sense of a sentence in which a number word occurs." But what is it to explain the sense—or *meaning*, to abide by our terminological convention—of a sentence? What sort of explanation does Frege have in mind? This question is addressed in the next few sentences of *Grundlagen*. In the first sentence Frege introduces the important notion of a *criterion of identity* and claims that all forms of reference make essential use of a criterion of identity:

If we are to use the symbol a to signify an object, we must have a criterion for deciding in all cases whether b is the same as a, even if it is not always in our power to apply this criterion. ((Frege, 1953), 73)

Note in particular that this claim about the need for criteria of identity is completely general and not restricted to natural numbers or even abstract objects.

Frege then continues as follows:

In our present case, we have to explain the meaning ['Sinn'] of the sentence 'the number which belongs to the concept F is the same as that which belongs to the concept G'; that is to say, we must reproduce the content of this sentence in other terms, avoiding the use of the expression 'the number which belongs to the concept F'.

This passage provides important information about the sort of explanation of meaning that Frege has in mind. He talks about explaining the meaning of a sentence, which he then glosses as follows: "that is to say, we must reproduce the content of this sentence in other terms." So Frege is here assuming that to explain the meaning of a sentence is to reproduce the content or meaning of the sentence in other (and presumably less problematic) terms. This understanding of the kind of explanation of meaning that the context principle calls for has been adopted by most of Frege's followers as well. A central theme of this article will be that this understanding of the kind of explanation that is needed is misguided, and that progress would be better served by adopting a very different understanding, which will be spelled out Section 6. The passage also imposes a requirement on the criterion of identity to be provided for natural numbers: When we explain the meaning of a basic numerical identity, we must avoid any use of the terms that flank the identity sign. Given the understanding of the explanatory task in question—to reproduce the content of basic numerical identities in other terms this seems like a reasonable requirement. For if we made use of the terms in question or of any co-referring terms, then the resulting explanation of the meanings of basic numerical identities would be as problematic as these identities themselves, and there would thus be no explanatory progress. This requirement can be formulated as follows.

Elimination Requirement: When explaining the meaning of a basic numerical identity, we must not make use of the singular terms that flank the identity sign or of any co-referring terms.

How can this requirement be satisfied? Frege proposes an ingenious answer based on a principle which he attributes to Hume.¹³ According to this principle, which is now known as *Hume's Principle*, the number of Fs is identical to the number of Gs if and only if the Fs and the Gs can be one-to-one correlated. The principle can be formalized as

(HP)
$$\#F = \#G \leftrightarrow F \approx G,$$

where '#F' means the number of Fs, and where $F \approx G$ is some second-order formalization of the claim that there is a relation that one-to-one correlates the Fs and the Gs.¹⁴ Frege proposes that instances of the right-hand side of (HP) can be used as an explanation of the meaning of corresponding instances of the left-hand side. This proposal clearly satisfies the Elimination Requirement, as the right-hand side of (HP) does not use any of the terms that flank the identity sign on the left-hand side or any co-referring terms.

What is controversial about Frege's proposal is the semantic claim that instances of the right-hand side of (HP) reproduce the meanings of the corresponding instances of the left-hand side. Frege is aware of this and goes on to defend the claim at some length. This is done in terms of an analogous but slightly simpler case, namely that of directions. Frege notes that

¹³Although the principle was only much later put to systematic use by people like Georg Cantor.

¹⁴Hume's Principle is now known to have an amazing mathematical property. Call the second-order theory with (HP) as its sole non-logical axiom *Frege Arithmetic*. Then a technical result known as *Frege's Theorem* says that Frege Arithmetic, along with some very natural definitions, allows us to derive all of second-order Peano Arithmetic. This result is hinted at in (Parsons, 1965) and explicitly stated and discussed in (Wright, 1983). For a nice proof, see (Boolos, 1990).

two directions are identical just in case the lines whose directions they are, are parallel. We thus have the following criterion of identity for directions:

(D)
$$d(a) = d(b) \leftrightarrow a \parallel b$$

where a and b range over directions or other directed items. Again it is uncontroversial that this proposal satisfies the Elimination Requirement. And again Frege makes the controversial claim that instances of the right-hand side reproduce the meanings of the corresponding instances of the left-hand side. He defends this latter claim as follows.

Thus we replace the symbol \parallel by the more generic symbol =, through removing what is specific in the content of the former and dividing it between a and b. We carve up the content in a way different from the original way, and this yields us a new concept. ((Frege, 1953), §64)

Here the principle of Sentential Determination is clearly at work. The meaning given by an instance of the right-hand side and the fact that the corresponding instance of the left-hand side has the logical form of an identity statement determine the meanings of the direction-terms that are being introduced. The defensibility of this answer will occupy us in the next two sections.

More generally, (HP) and (D) belong to a class of principles known as *abstraction principles*, which have the form

(*)
$$f(\alpha) = f(\beta) \leftrightarrow \alpha \sim \beta,$$

where α and β are variables, \sim is an equivalence relation on the kind of entities that α and β range over, and f is a function from such entities to objects. In all these cases the Elimination Requirement is clearly satisfied. The natural generalization of Frege's other claim is that matching instances of the two sides of an abstraction principle are—at least in suitable cases—just different carvings up of a shared meaning. This claim can be formulated as follows.

Recarving Thesis: Matching instances of the two sides of an abstraction prin-

ciple have the same meaning.

The precise content of the thesis will obviously depend on how the notion of meaning is understood. For present purposes it suffices to observe that the notion of meaning must be the same as that which figures in the context principle and Frege's Proposal.

The explanatory potential of the Recarving Thesis is huge. If correct, the thesis will provide a strategy for explaining how reference to numbers and other abstract objects is possible. For if sentences involving such reference are synonymous to, or recarve the meanings of, sentences that involve no such reference, then the latter sentences will provide an unproblematic semantic and epistemic handle on the former. It is thus not surprising that Frege and many of his followers have been enticed by this approach to the problem of reference.¹⁵

However, the approach also faces a variety of challenges. One of these challenges—the socalled "Julius Caesar problem"—made Frege back off from this approach and instead adopt what would now be regarded as set-theoretic definitions of most mathematical objects. I will not have anything to say about the Caesar problem here. My focus will rather be on some problems posed by the Recarving Thesis.¹⁶ In the next section I argue that this thesis conflicts with a strong form of the principle of compositionality, and in Section 5, that the thesis sits poorly with Frege's Proposal itself. I then go on to explain how Frege's Proposal can and should be developed in a way that eliminates any need for the Recarving Thesis.

4 The Recarving Thesis and Strong Compositionality

How plausible is the claim that instances of the left-hand side of a suitable abstraction principle are identical in meaning to the corresponding instances of the right-hand side? Consider the case of directions. If instances of the left-hand side of (D) really are identity statements involving singular terms referring to directions, how can their meanings be given by the corresponding instances of the right-hand side, which involve no such reference? Must not basic semantic features of a sentence—such as what objects it refers to—be preserved by any adequate characterization of its meaning? This worry is widely shared. For instance, the nominalist Hartry Field argues that since instances of the left-hand side of (D) purports to refer to abstract objects whereas instances of the right-hand side do not, the Recarving Thesis cannot possibly be true.¹⁷ The anti-nominalist Michael Dummett concurs, arguing

¹⁵The Recarving Thesis has been endorsed not just in Frege's *Grundlagen* but also by Bob Hale and Crispin Wright in (Wright, 1983) and in many of the articles collected in (Hale and Wright, 2001), esp. (Hale, 1997) but also essay 5, pp. 149-150; essay 8, pp. 192-197; and essay 12, pp. 277-278. However, it should be kept in mind that the use to which Hale and Wright put the Recarving Thesis is not always the same as that which I have attributed to Frege.

 $^{^{16}}$ Some other problems and challenges will be summarized in Sections 7.3 and 7.4 below.

 $^{^{17}\}mathrm{See}$ (Field, 1984).

that for instances of the two sides to be sufficiently close in meaning, the singular terms on the left-hand side cannot possibly have genuine reference but must be "semantically inert."¹⁸ Field and Dummett are here both relying on the following thesis:

Meanings Involves Referents: Any characterization of the meaning of a sentence S that contains a referential occurrence of a singular term a must make use of a itself or some co-referring term.

Now, one way to respond to Field and Dummett would be by giving up what has so far been an implicit background assumption, namely the following principle.

Surface Syntax: The two sides of an abstraction principle have the syntactic and semantic form that they appear to have.

If one is willing to give up this principle, two cheap ways of holding on to the Recarving Thesis become available. One option is to regard the right-hand side as just an unconventional way of writing the left-hand side. But this is unattractive because the right-hand side would then become just as problematic as what we set out to explain. We wanted to explain how reference to a particular sort of abstract object is possible. But on this approach the explanation offered would proceed via a sentence that makes use of precisely the kind of reference in question. Another option is to regard the left-hand side as just an unconventional way of writing the right-hand side. But this option too is unattractive. For on this account the left-hand side won't involve any genuine reference to directions or other abstract objects but only the (misleading) syntactic appearance of such reference.¹⁹ So adopting this option would simply be to give up on our original explanatory task. Because of these considerations, I will in what follows hold on to Surface Syntax as a background assumption.

Given this assumption, there is a direct conflict between the Recarving Thesis and the thesis that Meaning Involves Referents. For instances of the left-hand side of an abstraction principle contain singular terms that are not found in the corresponding instances of the right-hand side. And by Surface Syntax, this observation reflects the actual syntactic and semantic forms of the relevant instances. It then follows from Meaning Involves Referents that instances of the right-hand side cannot have the same meaning as the corresponding instances of the left-hand side. But this contradicts the Recarving Thesis, which says that matching instances of the two sides do have the same meaning.

¹⁸See (Dummett, 1991b), esp. ch. 15.

¹⁹A sophisticated version of this response is developed in (Dummett, 1991b), especially chapter 15.

4.1 The problem on Frege's mature theory of sense and reference

I will now examine whether it is possible to hold on to the Recarving Thesis by rejecting the thesis that Meaning Involves Referents. In order to do so, we need to replace the informal notion of meaning we have employed thus far with a more developed notion. I begin by discussing the problem as it arises in the context of Frege's mature theory of sense and reference. This theory is not only of obvious relevance to any study of Fregean ideas; it is also one of the more promising theories of meaning on offer.

At the heart of Frege's theory is a theory of reference, which ascribes *semantic values* (or *Bedeutungen*) to expressions of all syntactic categories.²⁰ I will write $[\![E]\!]$ for the semantic value of the expression **E**. Frege argues that the semantic value of a sentence is its truth-value and that the semantic values of other expressions are their contributions to the truth-values of sentences in which they occur. Moreover, he argues that these semantic value of a complex expression is determined as a function of the semantic values of its individual sub-expressions. For instance, the semantic value of an atomic sentence $\mathbf{P}(\mathbf{a}_1, \ldots, \mathbf{a}_n)$ is functionally determined in accordance with the following equation:

(1)
$$\llbracket \mathbf{P}(\mathbf{a_1}, \dots, \mathbf{a_n}) \rrbracket = \llbracket \mathbf{P} \rrbracket (\llbracket \mathbf{a_1} \rrbracket, \dots, \llbracket \mathbf{a_n} \rrbracket).$$

Having explained the notion of semantic value (or *Bedeutung*), Frege goes on to explain the *sense* of an expression as *the mode of presentation* of its semantic value. Since this mode of presentation includes the way in which the semantic value of the complex expression is determined as a function of the semantic values of its simple constituents, the sense of an expression will inherit from the underlying semantic theory a certain compositionality. In particular, the sense of a sentence is given by its *canonical* truth-condition, which displays how its truth-value depends functionally on the semantic values of all of its simple constituents. We thus get the following thesis.

Truth-conditions Involve Referents: The canonical truth-condition of a sentence \mathbf{S} involves the referents of all singular terms that occur in referential positions in \mathbf{S} , in the sense that names of these referents must occur in referential positions in the characterization of \mathbf{S} 's canonical truth-condition in the meta-language.

²⁰The reading of Frege I am about to present relies heavily on the work of Michael Dummett. See e.g. (Dummett, 1978), chapter 9 and (Dummett, 1981a), chapter 5.

But this thesis is just a special case of the more general thesis that Meanings Involve Referents. So this thesis too conflicts with the Recarving Thesis.²¹

A particularly interesting case are identity statements. Applying the above thesis to a simple identity statement $\lceil \mathbf{a} = \mathbf{b} \rceil$, it follows that its canonical truth-condition must directly involve the referents $\llbracket \mathbf{a} \rrbracket$ and $\llbracket \mathbf{b} \rrbracket$. This means that the truth-condition must have the form $R(\llbracket \mathbf{a} \rrbracket, \llbracket \mathbf{b} \rrbracket)$ for some dyadic relation R. Since the instances of the right-hand side of an abstraction principle do not have this form, they cannot give the sense of the identity statements which are instances of the left-hand side. In fact, it is hard to see how the relation R could be anything other than the relation of identity. This gives rise to the following canonical truth-condition:

(2)
$$\lceil \mathbf{a} = \mathbf{b} \rceil \text{ is true iff } \llbracket \mathbf{a} \rrbracket = \llbracket \mathbf{b} \rrbracket.$$

But this canonical truth-condition, which does give the sense of the identity statement on the left-hand side, refers to precisely the kind of object in question and thus violates the Elimination Requirement from the previous section. I conclude that on Frege's mature theory of sense and reference, the Recarving Thesis is false of the notion of sense.

4.2 The conflict with Strong Compositionality

Are there other notions of meaning that are more hospitable to the Recarving Thesis? We have seen that this would have to be a notion of meaning for which the thesis that Meaning Involves Referents is false. In order to chart the possible options, I will now examine what other adjustments would be required in order to hold on to the Recarving Thesis.

Let's begin by considering the principle of compositionality, which can be given the following formulation.

Compositionality: The meaning of a complex expression is determined by the meanings of its constituent parts, in accordance with their syntactic combination.

Consider a simple subject-predicate sentence like 'John runs'. According to Compositionality, the syntactic operation of predication is associated with an operation on semantic values. On many accounts, this operation will simply be that of function-application. The semantic value

 $^{^{21}\}mathrm{Recall}$ that Surface Syntax has been adopted as an implicit background assumption.

of our simple sentence will then be determined by the equation:

$$[3] \qquad [John runs] = [runs]([John])$$

(which of course is just a special case of (1) above). That is, the semantic value of the sentence 'John runs' is the result of applying the function which is the semantic value of the predicate 'runs' to the argument which is the semantic value of the subject 'John'.

The principle of compositionality in the above form remains compatible with the Recarving Thesis. For although this principle says that the meaning of a sentence *can* be explained in terms of the meanings of its constituents, it doesn't say that the meaning of the sentence *must* be so explained. This removes any push towards the thesis that Meanings Involve Referents.

However, we will see shortly that many adherents of compositionality are also committed to the following, much stronger principle:

Strong Compositionality: The meaning of a complex expression is built up from the meanings of its constituent parts, in accordance with their syntactic combination.

This principle ensures that any adequate characterization of the meaning of a sentence *must* involve the meanings of its constituents.²² Strong Compositionality therefore conflicts with the Recarving Thesis. To establish this, we first use our background assumption of Surface Syntax to show that matching instances of the two sides of an abstraction principle have different syntactic and semantic structure: for the instance of the left-hand side involves singular terms not found in the instance of the right-hand side. It then follows by Strong Compositionality that the two instances must differ in meaning. But this conflicts with the Recarving Thesis, which says that matching instances of the two sides have the same meaning.

To what extent should this worry adherents of the Recarving Thesis? The answer obviously depends on the plausibility of Strong Compositionality. The considerations of the previous subsection show that Strong Compositionality is built into truth-conditional theories of meaning. It is therefore not surprising that Frege repeatedly and explicitly commits himself to Strong Compositionality concerning sense.²³ Strong Compositionality is also an

 $^{^{22}}$ Note also that Strong Compositionality entails the principle of Sentential Determination from Section 2, that is, the version of the context principle which says that the meaning of a sentence determines the meanings of its constituents.

²³See for instance (Frege, 1964) and (Frege, 1963), both of which will be discussed in Section 6.1 below.

essential part of Russellian theories of meaning, which regard the meaning of a sentence as a structured proposition, composed of worldly items corresponding to the various meaningbearing parts of the sentence. However, there are also various coarser notions of meaning for which Strong Compositionality does not hold. For instance, if meaning is taken to be just Fregean reference, then Strong Compositionality obviously fails. For the reference of a sentence is just its truth-value, which is not built up from the referents of the sentence's constituent parts. Strong Compositionality also fails on many possible worlds accounts of meaning. For instance, if the meaning of 'John runs' is a set of worlds, then this meaning will not (in any appropriate way) be built up from the meanings of its constituents. The significance of these non-strongly-compositional notions of meaning for the present debate will be discussed in the next section.

Let's take stock of this section. I first argued that we have good reason to hold on to Surface Syntax as a background assumption. Then I observed that the Recarving Thesis conflicts with the thesis that Meanings Involve Referents, and I showed that this latter thesis is implicit in Frege-inspired truth-conditional theories of meaning. This is obviously disastrous for anyone attracted *both* to the strategy of *Grundlagen* for explaining reference *and* to Frege's mature theory of meaning. Nevertheless, the option remains of defending the Recarving Thesis by adopting some alternative notion of meaning which avoids the claim that Meanings Involve Referents. In order to explore the prospects for this strategy, I showed that Strong Compositionality—which is a very general principle that is shared by a great variety of theories of meaning—still conflicts with the Recarving Principle. My conclusion is thus that any defense of the Recarving Principle will have to rely on a notion of meaning for which Strong Compositionality fails.

5 Meaning without Strong Compositionality?

This conclusion is not new. In fact, most philosophers with a strong commitment to the Recarving Thesis have probably realized that they need a notion of meaning which isn't strongly compositional.²⁴ This is almost certainly true of Frege himself. Admittedly, when the Recarving Thesis was first advanced in *Grundlagen*, Frege may not yet have been aware of the compositionality principles for meaning. But had the principle of Strong Compositionality

 $^{^{24}}$ See (Hale, 1997), which is the most developed defense to date of the Recarving Thesis, but also (Hale and Wright, 2001), pp. 192-197. See also (Heck and May,).

been put to him, he would almost certainly have rejected it, realizing that it conflicts with his claims about our ability to "recarve content."

Moreover, it is likely that Frege in *Grundlagen* was to some extent drawing on a theory of meaning that was hinted at, but never properly developed, in his first book *Begriffsschrift* from 1879. Here Frege characterizes the "conceptual content" of a sentence in terms its inferential relations to other sentences. When two sentences or judgments are equivalent in that

the conclusions that can be drawn from one when combined with certain others also always follow from the second when combined with the same judgements [...] I call that part of the content that is the *same* in both the *conceptual content*. Since *only this* has significance for the *Begriffsschrift*, no distinction is needed between sentences that have the same conceptual content. ((Frege, 1879), p. 53)

So if by 'meaning' we mean this notion of conceptual content, then Strong Compositionality will almost certainly fail.

What exactly would be required of a non-strongly-compositional notion of meaning in order for it to validate the Recarving Thesis and serve the needs of the associated explanatory project? The notion of meaning would obviously have to be sufficiently coarse-grained to allow matching instances of appropriate abstraction principles to have the same meaning. On its own this requirement is not hard to satisfy. For a silly example, let the meaning of every (simple or complex) expression be the Roman emperor Julius Caesar. Then instances of both sides of an abstraction principle will trivially have the same meaning. For a more interesting example, let the meaning of an expression be its Fregean reference. On the assumption that the relevant abstraction principles are (or can be taken to be) true, the Recarving Thesis will then be satisfied. But Frege's dealings with the Recarving Thesis strongly suggest that he had in mind a much more fine-grained notion of meaning. It seems that instances of the two sides of appropriate abstraction principles were supposed to be equivalent in a much tighter way than just material equivalence, as would be the case if the notion of meaning at work was just that of Fregean reference. Let's therefore examine what further constraints can be placed on any notion of meaning capable of figuring in the explanatory project associated with the Recarving Thesis.

One constraint is that the notion of meaning must be well motivated and do substantive theoretical work. Regardless of precisely how this constraint is spelled out, it will no doubt rule out the silly example above.

A second constraint is that the notion of meaning must leave some privileged role for abstraction principles. For Frege clearly thinks that abstraction principles are particularly well suited for the recarving of meanings and the conferral of reference on otherwise problematic singular terms. This view is also shared by most of his followers, such as the neo-Fregeans Bob Hale and Crispin Wright, whose project is based in very large part on abstraction principles. This privileged role of abstraction principles must be clearly reflected in the notion of meaning that is employed. More precisely, this notion must identify the meanings of instances of the two sides of suitable abstraction principles but not identify the meanings of too many other pairs of sentences, in particular not the meanings of pairs of sentences to which one is unwilling to apply Frege's strategy. For instance, the meanings of the Dedekind-Peano axioms must not be identified with those of any logical truths, as there would otherwise be a route to the axioms of arithmetic which is simpler and more direct than the more laborious ones developed by various kinds of logicists.²⁵

A third constraint is that the notion of meaning must be one which makes plausible the versions of the context principle that the argument relies on. In particular, the meaning of a sentence must be such that it determines the meanings of its constituents, as required by Sentential Determination. This will require that the notion of meaning be quite rich. For instance, the Fregean reference of a sentence is far too impoverished to determine the reference of the various constituents of the sentence. For given only the truth-value of a sentence, there is no way to retrieve the referents of its constituents. The same goes for content construed as a set of possible worlds. For given only a set of possible worlds which is assigned to a sentence as its content, we cannot in general retrieve the referents of any of the constituents of the sentence.

Is there some notion of meaning that satisfies all of these constraints? The most sophisticated attempt to date to articulate such a notion of meaning is due to Bob Hale.²⁶ His proposal can be regarded as a refinement of Frege's idea from *Begriffsschrift*, where "the conceptual content" of a sentence was said to be determined by its inferential relations to other sentences. The key question is what sort of entailments we should be concerned with. Hale proposes that the relevant notion of entailment is a broadly logical one (which includes

 $^{^{25}}$ (MacFarlane, 2008) challenges Hale and Wright to explain what is so special about abstraction principles. Much of the answer given in (Hale and Wright, 2008) has to do with abstraction principles' being uniquely well suited to recarve meanings.

 $^{^{26}}$ See in particular (Hale, 1997).

analytic entailment) but which is restricted so as to require that the entailment be *compact*, roughly in the sense that the premises contain no redundant content that is inessential to the entailment.²⁷ His guiding idea is then that

Two sentences have the same [meaning] iff anyone who understands both of them can tell, without determining their truth-values individually, and by reasoning involving only compact entailments, that they have the same truth-value. (*ibid.*)

This guiding idea is then used to motivate a precise definition of sameness of meaning.²⁸

Hale's notion of meaning has been designed precisely so as to validate the Recarving Thesis. It does this because the two sides of an acceptable abstraction principle compactly entail each other. But what about the three further constraints that we identified? There is some reason to worry about the first constraint. Is there really room for another notion of meaning, intermediate between sense and reference? This new notion of meaning would presumably have to be integrated into existing semantics in a natural and well-motivated way. This would be an enormous undertaking, work on which has not even begun. But since this constraint has not been given a very precise formulation, I will let the worry pass: worse is anyway to come.

What about the second constraint—that the notion of meaning leave a privileged role for abstraction principles? An objection due to Michael Potter and Timothy Smiley is relevant here. They show that on Hale's notion of meaning all true arithmetical identities—such as '7 + 5 = 12' and $'2^3 = 8'$ —have the same meaning as '0 = 0' (and thus by transitivity also the same meaning as each other).²⁹ This is bad news for Hale. It appears that on his notion of meaning a vast range of arithmetical truths can be obtained by recarving the meaning of a simple identity, without the slightest need for abstraction principles. However, it should be mentioned that Potter and Smiley's trivialization result is highly sensitive to the details of Hale's proposal. It is thus not impossible that the proposal can be patched up in a way that prevents this kind of trivialization.

The most problematic constraint is the third one, which requires that the meanings of instances of the right-hand side of an acceptable abstraction principle suffice to determine

 $^{^{27}}$ The official definition of compactness has undergone some refinement over the years. The latest version is found at (Hale and Wright, 2001), p. 112 and is reproduced as Definition 1 in an appendix to this article.

 $^{^{28}}$ The definition is given at (Hale and Wright, 2001), pp. 113-115 and is reproduced as Definition 2 in the appendix.

²⁹See (Potter and Smiley, 2002) but also (Potter and Smiley, 2001). This argument is reproduced in the appendix.

the meanings the singular terms introduced in the corresponding instances of the left-hand side. This gives rise to an objection due to Kit Fine, which applies not just to Hale's proposal but much more broadly.³⁰ According to this objection, no notion of meaning that is coarsegrained enough to validate the Recarving Thesis can effect this kind of determination. Let \Leftrightarrow be the relation of sameness of meaning. For a concept F, let F^+ be the concept that is true of all and only the Fs and the smallest natural number that is not an F. Consider two concepts F and G. According to the Recarving Thesis we have the following two relations:³¹

(4)
$$\#F = \#G \Leftrightarrow F \approx G,$$

(5)
$$\#F^+ = \#G^+ \Leftrightarrow F^+ \approx G^+,$$

But according to Hale's definition of sameness of meaning, we also have:

(6)
$$F \approx G \Leftrightarrow F^+ \approx G^+.$$

By transitivity of the relation \Leftrightarrow it thus follows that:³²

(7)
$$\#F = \#G \Leftrightarrow \#F^+ = \#G^+.$$

This means that the independently given meaning with which we began—that of ' $F \approx G$ ' can be recarved in two completely different ways: either as #F = #G or as $\#F^+ = \#G^+$. But these latter two sentences involve reference to distinct pairs of numbers. Our notion of meaning is therefore too coarse-grained to enable the meaning of an instance of the right-hand side to uniquely determine the meaning of the corresponding instance of the left-hand side.

Note also that this objection has force not just against Hale's notion of meaning but much more broadly. For it is hard to see how a notion of meaning that is coarse enough to validate (4) and (5) could fails to validate (6) as well.

At the root of this objection and the previous one lies in the following problem. On the one hand we need a notion of meaning that validates the Recarving Thesis. As we saw in the previous section, this will have to be a notion of meaning for which Strong Compositionality fails. On the other hand we would like Sentential Determination to hold and the meaning of

³⁰See (Fine, 2002), pp. 39-41.

³¹For increased readability I drop quotation marks around the sentences flanking the ' \Leftrightarrow '.

 $^{^{32}\}mathrm{The}$ argument is spelled out in the appendix.

a sentence thus uniquely to determine the meanings of its constituents expressions. But for this to hold, the meaning of a sentence must already implicitly contain the meanings of its constituent expressions. But this pushes us back towards Strong Compositionality. There is thus good reason to believe that the project of *Grundlagen*, with its essential commitment to both the Recarving Thesis and Sentential Determination, is doomed.

6 The context principle without the Recarving Thesis

Somewhat surprisingly, this may have been Frege's considered view as well. For as I will now go on to explain, there is evidence that Frege continued to regard the context principle as essential to any account of reference even after he had developed his theory of sense and reference and explicitly committed himself to Strong Compositionality for senses and the thesis that Truth-conditions Involve Referents. This means that Frege remained committed to what I have called "Frege's Proposal" even after he adopted a view of meaning which quite clearly is incompatible with the Recarving Thesis. And this suggests the surprising conclusion that the context principle and its use in the explanation of reference may not need the Recarving Thesis after all, contrary to the view of the *Grundlagen* and that of most of Frege's followers.

6.1 The context principle after *Grundlagen*

In *Grundlagen*, where Frege talks bout "recarving content," there is no commitment to any of the theses which we have seen to conflict with the Recarving Thesis. But in later works Frege explicitly commits himself to two such theses, namely Strong Compositionality and the thesis that Truth-Conditions Involve Referents. For instance, *Grundgesetze* is committed to Strong Compositionality concerning senses:

If a name is part of the name of a truth-value [i.e. part of a sentence], then the sense of the former name is part of the thought expressed by the latter name [i.e. by the sentence]. ((Frege, 1964), §32)

So Frege here claims that when an expression is part of a sentence, then the sense of the former is part of the sense of the latter. Strong Compositionality plays a particularly important role in Frege's thought in the later years of his career, where it is often invoked to explain the socalled "productivity" of language, that is, our ability to produce and understand a potential infinity of different sentences. The following passage provides a good illustration.

It is astonishing what language can do. With a few syllables it can express an incalculable number of thoughts, so that even a thought grasped by a terrestrial being for the very first time can be put into a form of words which will be understood by someone to whom the thought is entirely new. This would be impossible, were we not able to distinguish parts in the thoughts corresponding to the parts of a sentence, so that the structure of the sentence serves as the image of the structure of the thoughts. ((Frege, 1963), p. 1)

It is hard to imagine that the author of these passages could still have been committed to the Recarving Thesis.

Does this mean that Frege gave up on the idea of using the context principle to explain reference? This view has been defended by Michael Resnik.³³ But this cannot be right. For *Grundgesetze* I §§29-31 makes very heavy use of a contextual account of what is required for a proper name to refer. The main change is that in *Grundgesetze* the category of sentences is subsumed under that of proper names. It is therefore no longer an option to say that a proper name refers provided that all sentential contexts in which the name occurs are meaningful; rather we have to consider all kinds of contexts. This is precisely what Frege does in the "generalized context principle" of *Grundgesetze*. This principle lays down the following requirement for a proper name to refer:³⁴

A proper name has denotation if the proper name that results from that proper name's filling the argument-places of a denoting name of a first-level function of one argument always has a denotation, and if [a corresponding requirement holds for names of functions of many arguments]. (p. 84)

The observations made in this subsection show that Frege must somehow have regarded the generalized context principle as compatible with Strong Compositionality concerning sense and the thesis that Truth-Conditions Involve Referents. The question is how.

³³See (Resnik, 1967). Some of Dummett's writings give the impression that he agrees; see e.g. (Dummett, 1981a), pp. 7 and 495. However, the relevant passages are more plausibly read as noting that as of the 1890s Frege can no longer regard *sentential* contexts as privileged because of his subsumption of the category of sentences under that of proper names. Despite this Dummett claims that Frege "retains the context principle [...] as far as can be done without distinguishing between sentences and other complex proper names" ((Dummett, 1981b), p. 409). See also (Dummett, 1981b), chapter 19; (Dummett, 1991b), esp. chapter 17; (Dummett, 1991a), pp. 229-233; and (Dummett, 1995).

 $^{^{34}\}mathrm{See}$ (Heck, 1997) and (Linnebo, 2004) for discussion.

6.2 Semantics versus meta-semantics

Let's return to the work that the context principle is supposed to do in Frege's philosophy of mathematics. I argued in Section 2 that the most important aspect of the context principle is the thesis of Sentential Priority, which says that the reference of a singular term has to be explained in terms of an explanation of the meanings of certain complete sentences. But it is not clear *what sort* of explanation of meaning is called for. There are two main possibilities. One possibility is that we need to explain the meanings of the relevant sentences in other (and presumably simpler) terms. This is how the explanatory demand is understood in *Grundlagen*. Another possibility is that we need to explain what is responsible for these sentences' having the meanings that they have; that is, to explain what these semantic facts consists in.

This ambiguity in the phrase 'explanation of meaning' corresponds to a distinction between what we may call *semantics* and *meta-semantics*.³⁵ *Semantics* is concerned with how the semantic value of a complex expression depends upon the semantic values of its various simple constituents. We ascribe to each complex expression a certain semantic structure and explain how its semantic value is determined by this structure as a function of the semantic values of its simple constituents. The various compositionality principles thus belong to semantics. *Meta-semantics*, on the other hand, is concerned with what is involved in an expression's having the semantic properties that it happens to have, such as its semantic structure and its semantic value.

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. Metaeconomics, 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 euros. 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.

With the distinction between semantics and meta-semantics on board, I claim that the explanation of reference that Frege seeks in *Grundlagen* §62 is meta-semantic, not semantic.

³⁵See (Stalnaker, 1997). My distinction between semantics and meta-semantics is thus the same as Stalnaker's distinction between "descriptive" and "foundational" semantics. See e.g. (Stalnaker, 2001).

The semantic question would be *which objects* various number terms refer to. But such questions have very simple answers—at least for someone like Frege who is both a platonist and unconcerned about the indeterminacy of reference. For instance, the decimal numeral '7' and the Roman numeral 'VII' both refer to the natural number 7. The question that Frege is concerned with is rather the meta-semantic one of what the relation of reference *consists in*. How can the numbers "be given to us," given that we cannot have any ideas or intuitions of them? In order to answer this question Frege proposes that we invoke the context principle. This will allow the question to be answered in terms of an explanation of the meanings of complete sentences in which the relevant term occurs.

Given the meta-semantic nature of the question how reference to natural numbers is possible, one would think that the required explanation of sentence meaning should be meta-semantic as well. That is, one would think that Frege's strategy is to transform a meta-semantic question concerning reference to another meta-semantic question concerning sentence meanings. But this is not how the required explanation of sentence meaning is understood in *Grundlagen*. For as we have seen, Frege there says that what is required as to "reproduce the content of this sentence in other terms, avoiding the use of the expression [whose reference is to be explained]" ((Frege, 1953), p. 73). So Frege here takes the required explanation of sentence meaning to be semantic, not meta-semantic. And as we saw in Section 3, it is solely in order to provide such a semantic explanation that Frege is led to the Recarving Thesis. But perhaps this understanding of the required explanation was a mistake all along. Perhaps everything should have been kept at the meta-semantic level. Then there would have been no need for the Recarving Thesis.

Let's investigate this hypothesis. I begin by formulating Frege's Proposal interpreted so as to keep everything at the meta-semantic level.

Frege's Proposal Refined: An explanation of what it is in virtue of which a singular term refers cannot be given entirely at the level of singular terms but must also involve an explanation of what it is in virtue of which associated identity statements mean what they do.

This is just like the original form of Frege's Proposal except that it is now made clear that all the explanations in question are meta-semantic. Although this proposal avoids any need for the problematic Recarving Thesis, it remains substantial. In fact, many of the leading attempts to give a meta-semantic account of the nature of reference would reject the proposal. For instance, a pure causal account or a historical-chain account would deny it.

Is the refined proposal compatible with Strong Compositionality? At first glance, there still appears to be a conflict. For according to the refined proposal, the reference of a singular term is be explained via an explanation of the meanings of the associated identity statements. So here the explanatory direction is top-down: by explaining the meanings of sentences we explain the reference of singular terms. But according to Strong Compositionality, the meaning of a complete sentence must be explained via an explanation of the meanings of its simple constituents, including its singular terms. So here the explanatory direction is bottomup: the meanings of subsentential expressions are used to explain the meanings of complete sentences.

However, this argument trades on an equivocation. For Strong Compositionality and Frege's Proposal Refined are concerned with completely different kinds of explanation of meaning. The former is concerned with a semantic explanation, whereas the latter is concerned with a meta-semantic explanation. This removes any direct conflict between the two principles.³⁶ Admittedly, semantics and meta-semantics still need to be integrated. But there is no reason why this cannot be done. Given the phenomenon of compositionality, a meta-semantic explanation of how a complex expression comes to be meaningful will obviously draw on meta-semantic explanations of how its various constituents come to be meaningful. What Frege proposes is just that some meta-semantic explanations at the most basic level will have to refer back to the sentential level; specifically, that the explanation of the reference of singular terms will have to refer back to identities involves these singular terms.

7 Towards an implementation of the refined proposal

According to the refined form of Frege's Proposal, the problem of explaining what it is for a singular term to refer should be transformed into the problem of explaining what it is for certain complete sentences involving this term to be meaningful. I have argued that this refined proposal avoids some of the most serious problems that threaten other attempts to

³⁶This reconciliation of the two principles is anticipated by an earlier proposal made at (Dummett, 1981a), pp. 3-7 and (Dummett, 1981b), p. 547. Dummett's proposal is that sentences are primary "in the order of explanation" of "what it is for sentences and words to have sense", whereas individual words are primary in "the order of recognition" of sense ((Dummett, 1981a), p. 4). Perhaps the main difference is that I, unlike Dummett, claim that the context principle applies in the first instance to the explanation of reference and only indirectly to the explanation of sense. I am also more explicit than Dummett is about the nature of the two principles and their different concerns.

explain reference by means of the context principle, in particular, that the refined proposal avoids the problematic Recarving Thesis and is fully compatible with Strong Compositionality. Even so, it is far from obvious that the refined proposal can successfully be carried out. Frege never provided much detail.

I will now outline a Frege-inspired theory of reference which is based on the refined proposal. Two caveats are needed. Firstly, although I believe Frege anticipated many aspects of this theory, he would probably have disagreed with other parts of it. But my goal in this section will be purely systematic and in no way exceptical. Secondly, I can here only attempt to provide a rough outline of the theory of reference. A more detailed account and defense will be given elsewhere. My present goal is merely to convey a sense of how the refined version of Frege's Proposal may be implemented.

7.1 The structure of a meta-semantic account of reference

My account will be based on the idea already encountered that reference has a rich and systematic structure. To begin with, objects are presented to us only via some of their parts or aspects. For instance, according to Frege, natural numbers are presented to us via concepts whose numbers they are, and directions are presented via lines or other directed items. I will refer to the items in terms of which an object is presented as *presentations*. Next, we understand how two presentations must be related for them to pick out the same object, namely that they must stand in some appropriate equivalence relation. Let's call these *unity relations*. Drawing on this structure, I propose that a singular term **a** refers in virtue of being associated with a presentation α and a unity relation \sim . Associated with \sim there is a function f such that the following criterion of identity holds:

(*)
$$f(\alpha) = f(\beta) \leftrightarrow \alpha \sim \beta$$

In particular we have $\llbracket \mathbf{a} \rrbracket = f(\alpha)$.

From a logical point of view the criterion of identity (*) is of course just an abstraction principle. However, the role that these principles play in the meta-semantic account of reference is very different from that which they play in *Grundlagen* and the work of the neo-Fregeans. In particular, on the meta-semantic account an abstraction principle is not supposed to effect any sort of recarving of meanings. The claim is rather that it is partially constitutive of the reference of a singular term \mathbf{a} that someone who understands the term takes the truth or falsity of identity statements involving \mathbf{a} to be a matter of the obtaining or not of the kinds of relations given by the right-hand side of the abstraction principle.

This account is an implementation of the refined form of Frege's Proposal. For an essential part of the explanation of what it is in virtue of which a singular term refers concerns the meaningfulness of certain identity statements. The explanation therefore enjoys the characteristic benefits associated with explanations based the context principle. In particular, the generality of the machinery of presentations and unity relations ensures that the explanation applies naturally to the forms of reference found in mathematics.

We may even attempt to extend the account to concrete objects. Consider for instance the case of ordinary physical bodies, that is, roughly, cohesive physical objects with natural boundaries. A physical body is most directly presented in perception, where we causally interact with one or more of its spatiotemporal 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.³⁷ However, for this extension to have any chance of success, we must allow the relation ~ on the right-hand side of (*) to fail to be reflexive (while remaining symmetric and transitive). For not every spatiotemporal part succeeds in determining a physical body. Correspondingly, the function f must be allowed to be partial, defined only on presentations α which are in the field of ~ (that is, defined only on presentations α such that $\alpha \sim \alpha$).

7.2 Restricting the scope of the account

However, the account just outlined is too ambitious. I will now describe some problems that the account faces and explain how these problems can be avoided by restricting the scope of the account.

We begin by noting that if the above account is applied directly to English or some other natural language, it will imply that every proper name is associated with a unique presentation and unity relation. But this conclusion is unacceptably strong. Every potential referent can be presented in countless different ways. For instance, the number 3 can be presented by means of any triply instantiated concept, and the direction North by any object pointing north. And there is absolutely no reason to think that these classes of presentations have

³⁷I elaborate on this view and defend it against some natural objections in my (Linnebo, 2005).

members that somehow play a privileged role.

This problem can be addressed by applying Frege's account first and foremost to thought rather than to language. This involves shifting attention from the problem of how *linguistic* expressions come to refer to the corresponding problem concerning mental representations. Our explanandum will thus be what is involved in someone's capacity for singular reference to various sorts of object. The Frege-inspired proposal that I will investigate is that this capacity for singular reference should be explained in terms of an explanation of what is involved in the person's capacity for understanding complete thoughts concerning objects of the sort in question.³⁸ This reformulation allows us to concentrate on an individual person rather than on a whole language community. And this in turn allows us to address the problem noted above. For we can now maintain that reference involves some Fregean mode of presentation while allowing this mode of presentation to vary with each individual act of reference. 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 the language.

We need a second restriction as well. This restriction is to focus on *canonical* cases of singular reference where the referent is "directly present" to the thinker.³⁹ 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. Other cases of canonical reference would be reference to a direction based on a perceived line or the reference to a shape based on a perceived figure. 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.

When these two restrictions are taken into account, we arrive at the following version of Frege's Proposal:

Frege's Proposal Refined and Restricted: An explanation of someone's ca-

 $^{^{38}}$ 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 natural; for in order to stand in some propositional attitude to a thought, one presumably needs to understand that thought.

³⁹In 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, 1981a), pp. 231-239 and (Evans, 1982), pp. 109-112.

pacity for canonical singular reference cannot be given entirely at the level of singular representations but must also involve an explanation of the agent's capacity for understanding identity statements concerning the object in question.

7.3 Are singular terms still rigid designators?

This meta-semantic account of reference faces a variety of other problems and challenges. I begin by explaining how we have the resources to address one of these problems, namely whether the account is compatible with the semantic thesis that singular terms are rigid designators. Consider a singular term **a** associated with a presentation α and a unity relation \sim . Let f be the function determined from \sim in accordance with the criterion of identity (*). I have argued that **a** refers, if at all, to the object $f(\alpha)$. One may then wonder whether my view collapses back into some version of the descriptivist view of names criticized by Kripke and others. Specifically, am I not committed to identifying the meaning of **a** with that of the description "the f of α ," with the result that **a** isn't a rigid designator after all? For instance, one concept can determine different natural numbers in different possible worlds.

Fortunately, we have the resources needed to address this worry. On my proposal, the nature of the function-argument structure $f(\alpha)$ is meta-semantic, not semantic. The expression **a** is semantically simple, and its semantic value (if any) is just the referent $f(\alpha)$. How this referent is determined is a meta-semantic matter, of no immediate semantic significance. As far as semantics is concerned, **a** is an atomic expression 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 structuring 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 $f(\alpha)$ is of non-semantic nature 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 $f(\alpha)$ 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.

7.4 Problems and challenges

I end by outlining three remaining problems, which will have to be addressed elsewhere.

First there is the so-called "bad company problem."⁴⁰ It is well know that many abstraction principles lead to paradox. The most famous example is Frege's Basic Law V, which says that two classes are identical just in case they are coextensive, or, in symbols:

(V)
$$\epsilon F = \epsilon G \leftrightarrow \forall x (Fx \leftrightarrow Gx).$$

For in second-order logic, (V) allows us to derive Russell's paradox. And other abstraction principles are unacceptable in more subtle ways. This calls for a principled demarcation of the abstraction principles that are acceptable from those that are not.

Next, is there really a function f of the sort that is needed to serve in an abstraction principle (*)? Even in non-paradoxical cases, it is far from obvious that such a function exists. There will for instance have to be a function that maps concepts to natural numbers. But nominalists, who deny that there are numbers, would also deny that any such function exists.

Finally, even if there is such a function f, will it be suitably unique? The problem is that we can prove that, if there is one candidate for playing the role of the function f, then there are many. For given any permutation of objects σ , the result $\sigma \circ f$ of composing f with σ will satisfy (*) just in case f satisfies (*). The question is thus whether we can point to some feature of the intended function f that distinguishes it from its unintended rivals.⁴¹

8 Conclusion

Let me summarize my main claims. Frege's context principle urges us to explain reference and other semantic properties in the context of complete sentences. In particular, the explanation of the reference of a singular term should proceed via an explanation of the meanings

 $^{^{40}}$ See (Linnebo, b) for a more extensive overview of the problem. I propose a solution to the problem in (Linnebo, a).

⁴¹See (Linnebo, c) for a proposed solution to these last two problems.

of identity statements in which the singular term occurs. But what is it to explain the meanings of these identity statements? The standard answer, which goes back to *Grundlagen* itself, has been that the required explanation is a restatement of the meanings in simpler and less problematic terms. This interpretation has led Frege and others to endorse the Recarving Thesis. But the Recarving Thesis is problematic because it conflicts both with Strong Compositionality and with Frege's Proposal that we use the context principle to account for reference.

Fortunately, as Frege himself later realized, the Recarving Thesis is not needed for the purposes of using the context principle to explain reference. The Recarving Thesis is introduced only as a result of a misunderstanding of the kind of explanation of sentential meaning required by the context principle. The relevant explanation should be meta-semantic, not semantic. When this is borne in mind, the conflict with Strong Compositionality disappears. This opens for an exciting Frege-inspired account of reference, which is applicable to mathematical as well as non-mathematical cases of reference.

This account obviously needs to be developed further. And it faces a number of problems and challenges, which will have to be addressed elsewhere. What I hope to have established in this article is that, when properly understood, the context principle may have an important role to play in the explanation of reference and that it is fully compatible with Strong Compositionality.

Appendix

I here provide some of the technical details that were omitted in Section 5. I begin with the definition of sameness of content from the postscript to the reprinting of (Hale, 1997).

Definition 1 (Compact entailment) Assume as given the notion of entailment. Then A_1, \ldots, A_n compactly entail B iff all of the following conditions are met:

- (i) A_1, \ldots, A_n entail B
- (ii) For any non-logical constituent E occurring in A_1, \ldots, A_n , there is some substitution E'/E such that the result A'_1, \ldots, A'_n of applying this substitution uniformly to A_1, \ldots, A_n does not entail B

(iii) For every subformula S of A, there is some formula S' which is modally equivalent to S—in the sense that both are necessary, both are contingent, or both are impossible⁴²— and which is such that the result A'_1, \ldots, A'_n of uniform application of the substitution S'/S to A_1, \ldots, A_n does not entail B

Definition 2 (Sameness of content) Let ' \Leftrightarrow ' stand for mutual compact entailment. Then two sentences S_1 and S_2 have the same content iff at least one of the following conditions is met:

- (i) $S_1 = S_2$
- (ii) $S_1 \Leftrightarrow S_2$
- (iii) There are sentences A and B such that $A \Leftrightarrow B$ and expressions E and E' such that S_1 and S_2 result from A and B respectively by uniform replacement of E by E'
- (iv) S_1 and S_2 stand in the transitive closure of the relation defined by the disjunction of (i), (ii), and (iii)

Claim 1 ((Potter and Smiley, 2002)) 6+4 = 10 has the same content as 0=0.

Proof. Let j be 'the number of planets'. Then by compact equivalence we have:

$$(8) j = 10 \Leftrightarrow j = 10 + 0$$

(9)
$$6+4 = 10+j \Leftrightarrow j = 0$$

Next clause (iii) of Definition 2 allows us to substitute '6+4' for j in (8) and '0' for j in (9) to get:

(10)
$$6+4 = 10 \Leftrightarrow 6+4 = 10+0$$

$$6+4 = 10+0 \Leftrightarrow 0 = 0$$

Claim 1 then follows by transitivity of \Leftrightarrow , which is secured by clause (iv) of Definition 2.

 $^{^{42}}$ Hale fails to specify how this is to be understood for open formulas. Presumably the notion of modal equivalence will have to be relativized to an assignment.

Claim 2 ((Fine, 2002)) Given the definitions of Section 5 we have:

(6)
$$F \approx G \Leftrightarrow F^+ \approx G^+$$

Proof. The two sides clearly mutually entail each other (in the relevant sense of entailment). To see that both entailments are compact we first note that every expression on either side can be substituted for in a way that destroys the entailment of the other side. (Here we use that F^+ and G^+ function as primitive expressions.) We also note that each side can be replaced by a sentence of the same modal status so as to destroy the entailment of the other side.

References

Beaney, M. (1997). The Frege Reader. Blackwell, Oxford.

- Benacerraf, P. (1973). Mathematical Truth. Journal of Philosophy, 70(19):661–679.
- Boolos, G. (1990). The Standard of Equality of Numbers. In Boolos, G., editor, Meaning and Method: Essays in Honor of Hilary Putnam. Harvard University Press, Cambridge, MA. Reprinted in (Boolos, 1998).
- Boolos, G. (1998). Logic, Logic, and Logic. Harvard University Press, Cambridge, MA.
- Dummett, M. (1956). Nominalism. *Philosophical Review*, LXV:491–505. repr. in (Dummett, 1978).
- Dummett, M. (1978). Truth and Other Enigmas. Harvard University Press, Cambridge, MA.
- Dummett, M. (1981a). Frege: Philosophy of Language. Harvard University Press, Cambridge, MA, second edition.
- Dummett, M. (1981b). *The Interpretation of Frege's Philosophy*. Harvard University Press, Cambridge, MA.
- Dummett, M. (1991a). Frege and Other Philosophers. Oxford University Press, Oxford.
- Dummett, M. (1991b). Frege: Philosophy of Mathematics. Harvard University Press, Cambridge, MA.

- Dummett, M. (1995). The Context Principle: Centre of Frege's Philosophy. In Max, I. and Stelzner, W., editors, Logik und Mathematik: Frege-Kolloquium Jena 1993. de Gruyter, Berlin.
- Evans, G. (1982). Varieties of Reference. Oxford University Press, Oxford.
- Field, H. (1984). Platonism for Cheap? Crispin Wright on Frege's Context Principle. Canadian Journal of Philosophy, 14:637–62. Reprinted in (Field, 1989).
- Field, H. (1989). Realism, Mathematics, and Modality. Blackwell, Oxford.
- Fine, K. (2002). The Limits of Abstraction. Oxford University Press, Oxford.
- Frege, G. (1879). Begriffsschrift: Eine der arithmetischen nachgebildete Formelsprache des reinen Denkens. Translated and reprinted in (Beaney, 1997).
- Frege, G. (1953). Foundations of Arithmetic. Blackwell, Oxford. Transl. by J.L. Austin.
- Frege, G. (1963). Compound Thoughts. Mind, 72:1–17. Originally published in 1923.
- Frege, G. (1964). Basic Laws of Arithmetic. University of California Press, Berkeley and Los Angeles. Ed. and transl. by Montgomery Furth.
- Frege, G. (1979). Posthumous Writings. Blackwell, Oxford.
- Hale, B. (1994). Dummett's Critique of Wrights Attempt to Resuscitate Frege. Philosophia Mathematica, 2:122–147. Reprinted in (Hale and Wright, 2001).
- Hale, B. (1997). Grundlagen §64. Proceedings of the Aristotelian Society, 97(3):243-61.
 Reprinted with a postscript in (Hale and Wright, 2001).
- Hale, B. and Wright, C. (2001). Reason's Proper Study. Clarendon, Oxford.
- Hale, B. and Wright, C. (2008). Focus Restored: Comments on John MacFarlane. Synthese. Forthcoming.
- Heck, R. G. (1997). Grundgesetze der Arithmetik I §§29-32. Notre Dame Journal of Formal Logic, 38(3):437–474.
- Heck, R. G. and May, R. The composition of thoughts.

Linnebo, Ø. Bad Company Tamed. forthcoming in Synthese.

- Linnebo, Ø. Introduction. forthcoming in Synthese.
- Linnebo, Ø. Meta-ontological minimalism.
- Linnebo, Ø. (2004). Frege's Proof of Referentiality. Notre Dame Journal of Formal Logic, 45(2):73–98.
- Linnebo, Ø. (2005). To Be Is to Be an F. Dialectica, 59(2):201–222.
- MacFarlane, J. (2008). Double Vision: Two Questions about the Neo-Fregean Program. Synthese. Forthcoming.
- Parsons, C. (1965). Frege's Theory of Number. In Black, M., editor, *Philosophy in America*. Cornell University Press. Reprinted in (Parsons, 1983).
- Parsons, C. (1983). Mathematics in Philosophy. Cornell University Press, Ithaca, NY.
- Pelletier, F. J. (2001). Did Frege Believe Frege's Principle? Journal of Logic, Language, and Information, 10:87–114.
- Potter, M. and Smiley, T. (2001). Abstraction by Recarving. *Proceedings of the Aristotelian* Society, 101:327–338.
- Potter, M. and Smiley, T. (2002). Recarving Content: Hale's Final Proposal. *Proceedings of the Aristotelian Society*, 102:301–304.
- Resnik, M. (1967). The Context Principle in Frege's Philosophy. Philosophy and Phenomenological Research, 27:356–365.
- Stalnaker, R. (1997). Reference and Necessity. In Hale, B. and Wright, C., editors, Blackwell Companion to the Philosophy of Language, pages 534–554. Blackwell, Oxford.
- Stalnaker, R. (2001). On Considering a Possible World as Actual. Proceedings of the Aristotelian Society, Suppl. vol. 65:141–156.
- Wright, C. (1983). Frege's Conception of Numbers as Objects. Aberdeen University Press, Aberdeen.