

Communication and strong compositionality

Peter Pagin

1. Cognitive and epistemic asymmetry

Suppose that you are given the task of translating ten words from English to Swedish. You know English and you don't know Swedish, but you have an English-Swedish dictionary. Given the dictionary, you easily accomplish what was asked for.

But then you are given a more difficult task. You are now to translate ten new words from Swedish to English, say 'låtsas', 'dimma', 'ofog', 'färdig', 'dumheter', 'glömska', 'om', 'springa', 'tryck', and 'tillbringare'. The reason this is more difficult is that you have to do it with the same tool as last time, the English-Swedish dictionary. The dictionary contains all translation pairs you need, but they are listed alphabetically according to the English members of the pairs. So you know where to find any English word in the list, but you don't know where to find the Swedish words (unless you already know the translation, which you don't). So in order to find the translation of 'ofog' you simply have to search through the dictionary, in whatever order you prefer, until you happen to find the word. You don't know how long it will take.

Because of this, the dictionary translation method is *cognitively asymmetric*. It defines a relation between the Swedish and English vocabularies, but it provides an efficient (e.g. as measured by average time needed) method only for one direction, not for the other. The hypothesis that a bilingual speaker's translation capacity is subserved by a mental English-Swedish dictionary can *explain* the speaker's ability of English-Swedish translation, but it cannot explain his ability of Swedish-English translation.

We can make things worse by two changes to this scenario. Suppose, first, that the list of words is infinite¹, and, second, that the strings of letters you are served need not be part of the vocabulary. In the English-to-Swedish task the increase in difficulty is not great. Given an arbitrary string of letters you know where to find it: if there is no such string at the determined position, it isn't a word in the English vocabulary. Hence, you have a decision method for determining wordhood, and what the translation is, if the word exists. By contrast, in the Swedish-to-English task, where you don't know where in the list to find the string, you only know that *if* there is a word, and you search systematically, you will sooner or later find it, even though not when. If the string (e.g. 'bab')

1. In order to get an enumeration the list is then ordered by length of the strings (longer strings later) and within each length alphabetically.

isn't in the vocabulary, however, your search doesn't come to an end; the idealized lexicon doesn't give a decision method for determining wordhood, i.e. whether a string is well-formed, and so no finite lexicon search will establish that the string isn't well-formed. Because of this, the idealized English-Swedish dictionary is also *epistemically asymmetric*. It offers an effective method for determining whether a string is part of idealized English, but not whether it is part of idealized Swedish.

Let's now take the mapping of simple expressions for granted, and consider asymmetry in methods of mapping syntactically complex expressions on other objects. We shall briefly look at three different examples, all of which involve a *compositional* mapping from some language L . The first is a translation mapping between L and a language L' , the last two are mappings between L and the natural numbers.

Example 1. Suppose that both L and L' are syntactically unambiguous (i.e. in neither case does the grammar produce one and the same surface structure in two different ways). Suppose that we have a translation function τ , with the following properties. First, the set A_L of simple expressions of L is mapped 1-1 and onto to the set $A_{L'}$ of simple expressions of L' . Second, the set \mathcal{O}_L of syntactic operations (rules) of L is mapped 1-1 and onto to the set $\mathcal{O}_{L'}$ of syntactic operations of L' . Third, for a complex expression (e_1, \dots, e_n) , formed by applying the operation α to the argument expressions (simple or complex) e_1, \dots, e_n , it holds that $\tau(\alpha(e_1, \dots, e_n)) = (\alpha)(\tau(e_1), \dots, \tau(e_n))$. So, τ is a homomorphism from L to L' , and it is easy to show that τ is an isomorphism as well. That is, τ is 1-1 and onto from L to L' , and its inverse τ^{-1} is a homomorphism from L' to L .

If we suppose that *parsing* is equally easy in both languages, it is as easy to translate by τ from L to L' as it is to translate by τ^{-1} from L' to L . Also, we do not have a well-formedness problem. For suppose that we can determine whether a string is a well-formed expression of L , and suppose that we have a string s that might or might not be a well-formed expression of L' . We can determine which by simply trying to apply τ^{-1} to s . If it can be applied, it delivers a well-formed expression of L as value, and then s is well-formed as well. Otherwise not.

τ is a cognitively symmetric translation method. The hypothesis that bilingual capacity is subserved by cognitive abilities that correspond to the elements of \mathcal{O} can explain translational capacity in both directions.

Example 2. We shall consider mappings from formulas of a formal language to natural numbers. Our first example is Gödel's own numbering, from strings over the vocabulary of a language L with class-typed variables, defined as follows (' f ' is the arithmetical successor operator):

'0'...1, 'f'...3, '~'...5, ' '...7, ' '...9, '('...11, ')'...13

If x is a variable of type n , then ' x '... p^n , where p is a prime number >13 .

For any string of these signs, with values n_1, \dots, n_k , the value of the string is $2^{n_1} \times 3^{n_2} \times \dots \times p_k^{n_k}$, where p_k is the k th prime number (Gödel 1931: 92).

Call this mapping ' G '. G is not epistemically asymmetric. There is an effective method for finding the prime number decomposition of any natural number n . Given a prime number decomposition of n , it is easy to see whether n is the value of a string over the alphabet, and if so, which.

However, G is cognitively asymmetric, for in general the prime number decomposition of a natural number is a considerably more complex computational task than that of computing the Gödel number of a sentence.

Example 3. Instead of Gödel's own method for mapping complex expressions on numbers, consider the following method F : for a particular string s , first compute the Gödel number $n=G(s)$, and then go the n :th position of the decimal expansion of π . Where the digit at the n :th position is k , find $G(e)$, where e is the $k+1$:th member of s (or the last member, if the length of s is k or smaller). Then $F(s)$ is the natural number denoted by the π expansion segment of length $G(e)$ and starting at position n .

This pretty contrived mapping is designed so that the F number belonging to a string is computable from the values of the elements of the string and the order in which they are arranged, and also so that there is no better method for finding out whether any string corresponds, by F , to a given natural number m —and if so, which—than going through the strings, computing their F numbers and comparing with m . If this is really the case (I find it likely, but abstain from trying to prove it), then F is not only cognitively asymmetric, but also epistemically asymmetric; there is no decision method in the direction from numbers to strings.

These examples will be used for comparison when considering methods for solving 'the communication problem'.

2. The communication problem

Let's assume a simplified picture of propositional linguistic communication. Speaker A has a thought, say the belief that p , and wants to tell hearer B that it is the case that p . I shall say that A successfully communicates to B that p , if and only if there is an utterance u such that

- a) A produces u because A intends to make B think that p ²

2. I think that a) is too strong, but the reasons for this are irrelevant in the present context.

b) B comes to think that p because of observing u .

In this context it is not relevant how we understand ‘because’, e.g. as meaning or implying the existence of a reason or intention, or only a causal connection, for instance between an intention to make someone think something and the effect of making an utterance. Choose your preferred account of communication intentions and reasons. What will matter here is the pairing of the utterance and the content.

I shall make the assumption, for which there is excellent empirical support, that linguistic communication virtually always succeeds, when it succeeds, because of the choice of the expression uttered. It is not, for instance, the case that success only depends on what the speaker thinks, not at all on what the speaker utters (which would be a kind of indirect telepathy). Rather, the choice of expression is of crucial importance, and that is an empirical fact, not something that defines what communication by means of utterances is.

Because we take the choice of expression as essential, we can restate the conditions of successful communication: A successfully communicates to B that p , if and only if there is an expression u such that

c) A utters u because A intends to make B think that p

d) B comes to think that p because of observing the utterance of u .

What are the conditions of success, given that telepathy is not an option? We can think of the interpreter B as having the task of finding the right interpretation in a domain of meanings, or possible interpretations. That domain is B ’s search space. We can assume that exactly one meaning—the proposition that p —is the right one, that is, the one intended by A . We can further assume that the search space is huge: the domain of possible interpretations is extremely big, maybe infinite. The interpreter’s task, so described, seems formidable.

Nonetheless, we are reliably successful, even if imperfect, in interpreting each other. What makes this success possible is that the interpreter is led from the expression to the meaning in a systematic way. If the expression has an established use for conveying a particular content (like *predator in the vicinity*) then success is explained in a fairly simple way, either as successful induction or, if the interpreter is more primitive, by some nomic connection between expression and interpretation. If the expression is new to the hearer, on the other hand, and he doesn’t have an innate disposition to interpret utterances of it, then the explanation must be of a different nature. It is customary to appeal to *the principle of compositionality*:

(PCF) The meaning of a complex expression is a function of the meanings of its parts and its mode of composition.

A standard story goes like this: the meaning of the whole is determined by the meanings of the parts and the mode of composition. Since the hearer knows the meanings of the simple parts, and knows the semantic significance of a finite number of syntactic modes of composition, and can parse the expression, i.e. recognize how it is built up out of simple parts, the interpreter can work out the meaning of the whole.³

Strictly speaking, that compositionality holds is not enough for making sure that an interpreter can figure out the meaning of a complex expression. It is not enough that the meaning of a complex expression is a function of the meaning of the parts and the mode of composition, for the function need not be computable. If it is not, the interpreter cannot work out what the value of the function is for new arguments. When compositionality is appealed to for explaining how one can understand new sentences, it is assumed, explicitly or tacitly, that the meaning of the whole depends on part meaning and mode of composition according to rules that are possible for the interpreter to follow, or proceed in accordance with, when interpreting new expressions. Let's call a meaning function μ , from well-formed expressions to meanings, *well-behaved* if it allows a speaker to effectively work out, in an intuitive sense, its values for arbitrary arguments.

But even with a well-behaved meaning function we *don't* have enough for explaining communication. For in order to explain communication we need to explain both how *B* finds the right interpretation, and how *A* finds an *appropriate linguistic item*, i.e. an expression that enables *B* to find the right interpretation. *A* has as his search space a huge domain of linguistic expressions, and needs to select one that will make *B*'s task solvable. The hearer solves a task of interpretation, and the speaker solves a task of expression, and we need to account for both.

The problem with the appeal to compositionality, even if it is an appeal to a well-behaved compositional meaning function, is that this function may be cognitively asymmetric. It may be that a meaning function μ that provides *B* with an efficient solution to the interpretation problem does not provide *A* with an efficient solution to the expression problem. If μ does not offer *A* any better method of selecting an expression than to search through the domain of syntactic items, and interpret them, one at a time, until he finds one whose interpretation coincides with the thought he wants to communicate, then we don't have an explanation of communicative success, for our actual performance is better than it would be by this method.

3. For reasons irrelevant to the present discussion, I think it is a mistake to appeal to knowledge of meaning. Rather, the account should be developed along the lines suggested at the end of this section.

To explain actual communicative success we would need a function that is cognitively *symmetric*. In terms of the examples of section 1, we would need a meaning function that is similar to the translation function of example 1, that is, a function μ by which something like the following *inverse* principle of compositionality would hold:

(IPCF) The expression of a complex content is a function of the expressions of its parts and its mode of composition.

If IPCF is true, then we can picture the production process in the following way: the speaker has a thought, say the belief that p , that he wants to convey to the hearer. He is aware of the structure of the proposition that p , and of its parts—call them ‘concepts’. He primitively associates the simple parts with linguistic items, and also the modes of composition by which the proposition is built with syntactic constructions. Putting the linguistic items together according to these syntactic constructions he arrives at a sentence that ‘expresses’ his belief, i.e. a sentence which the hearer can efficiently interpret as conveying the thought that p . If IPCF is true, we have an explanation, or rather the general outline of an explanation, of the capacity of natural language speakers of solving the problem of expression.

The IPCF hypothesis, as we can call it, will be investigated in the following section. Before pursuing this, however, I would like to consider an objection concerning explanatory value. It may be objected here that semantics falls strictly outside the explanation. For if we want to explain how a speaker can find the proper expression of a thought he is entertaining, it does not seem to matter how propositional contents, i.e. ideal, non-mental entities, are built up from parts, if they are at all. What matters seems to be the speaker’s mental *representation* of the propositional content. If the representation is a structured entity, then we can assume that the speaker has the ability of computing the linguistic expression (or his mental representation of the linguistic expression) from his representation of the thought content. This seems to be both necessary and sufficient. It is necessary, since if the propositional content itself is structured but its representation isn’t, then the speaker can’t compute the expression from his representation of content. And it is sufficient, since if the representation is appropriately structured, but the proposition isn’t, then the speaker nevertheless has the computational ability required.⁴

To the extent these objections are concerned strictly with issues of computation, i.e. transformations of mental representation, they are correct. However, if we are concerned with the explanation of successful *communication*, then a theory restricted to computa-

4. The sufficiency part of the objection is the gist of Stephen Schiffer’s argument, in Schiffer 1987, chapter 7, that public language semantics isn’t required for explaining communication. I have discussed Schiffer’s argument at length in Pagin *forthcoming*.

tion is insufficient. When communication succeeds, there is a thought content that is *shared* between speaker and hearer. Mental representations aren't shared. What is shared is the content of these representations. An explanation of why communication succeeds, or of how it succeeds, is incomplete if it doesn't consist in a theory that explains why the content of the speaker's representation is the same as the content of the hearer's representation. That is, a theory that stops after providing an account of the computational process itself, without considering content, doesn't provide an explanation of communicative success.⁵ It is not enough that the representations *in fact* have the same content; we need a theory that also predicts that they do.

On the other hand, in order to provide a full explanation of why communication succeeds, it is not enough to consider *only* the relation between syntax and semantics. For that doesn't explain how speaker and hearer manage to perform the transitions from the one to the other. However, what we *can* explain by considering syntax and semantics, while leaving mental states and processes out of the theory, is how it is *possible* for speaker and hearer to be so cognitively constituted that a communicative ability results. If syntax and semantics were only arbitrarily related, then no cognitive capacity short of telepathy (which we cannot explain) could make us good linguistic communicators. If, on the other hand, syntax and semantics, according to our theory, are so related that they *can* be combined with a plausible theory of mental representations and processes, then we have a theory that explains how it is possible that speaker and hearer together solve the communication problem.

More precisely, we have an explanation if syntax and semantics are so related, according to our syntactic theory *T* and our semantic theory *M*, that we can add a theory of mental representation *R*, such that

- i) *R* systematically assigns a representation of each element in the domain of *T* and *M* respectively, and
- ii) *R* systematically assigns a mental operation to each part of the definition in *M* of functions between the domains of *T* and *M*.

If our theories of syntax and semantics are such that a theory *R* *can* be added which meets these conditions, then we do have an account of how actual communicative success can be achieved. And for this we need only consider syntax and semantics itself. I shall give reasons for believing that natural language syntax and semantics does meet this condition.

5. This, in a nutshell, is the problem with Schiffer's account.

3. Bidirectional compositionality: the idea

There are two problems with IPCF that leap to the eye. The first is that in general we cannot speak of *the* expression of a particular thought, for often there are several—synonymous—expressions that can be used for expressing one and the same thought.⁶ However, this is a minor difficulty that will be handled by a technical adjustment below.

The second problem is substantial. IPCF presupposes that there are such things as complex contents, i.e. contents that have *proper parts* (or alternatively—if a complex content by definition is taken as the content of a complex expression—that complex contents do have proper parts). But this cannot just be taken for granted. Complex expressions are structured and they have meaning, but from this it does not follow that the meanings of complex expressions themselves are structured. What reason is there to think that they are?

Compare with the case for compositionality. First, there is one reason of a cognitive and epistemological nature to believe that natural languages are compositional, namely that this explains how hearers can know what a speaker means by the utterance of a sentence, even if both the sentence uttered and the content is new to the hearer. This reason has the form of an inference to the best explanation. To make it stick it has to be investigated whether there are any alternative explanations that are serious competitors. Such an investigation is non-trivial, but for the present purpose I'll just assume that some facts about communication do give us good reason to believe in the compositionality of natural language.

Second, if natural languages are compositional, then the following substitution version of compositionality holds:

(PCS) If the expressions e_1 and e_2 are synonymous, and the expression S' results from replacing e_1 by e_2 in S , then S and S' are synonymous.⁷

PCS is, under certain conditions to be specified below, equivalent with PCF. The immediate significance of this is that the hypothesis of natural language compositionality is open to fairly direct empirical testing. Take two expressions that are intuitively synonymous, and try out replacing the one by the other in various sentential contexts. If inter-substitution of synonymous expressions invariably preserves sentence meaning, in the cases investigated, then we have further empirical support for the belief in composition-

6. It may also be that two non-synonymous sentences that are both context-dependent, but in different ways, can be used for expressing the same content in some contexts but not in others, but I shall disregard this possibility. Of course, if for each content there is only one expression, then the meaning function is one-one.

7. It is assumed here that S' is meaningful if S is. See below.

ality of natural language. If in general it doesn't, then we have substantial disconfirmation of the hypothesis. The third possibility, i.e. the existence of a few types of exceptional counterexamples, or apparent counterexamples (like substitution in belief contexts) needs special consideration.⁸ For present purposes, let's assume that PCS in fact is valid for cases so far considered in English.

This gives us a dual support for the hypothesis: we have a reason from explanatory value, and we have further support from considering cases of substitution. Is there a parallel for the inverse compositionality that we have assumed? On the one hand, it was assumed that IPCF (relevantly qualified) does offer a potential explanation of how the speaker manages to select an appropriate linguistic expressions for conveying his attitudes to the hearer. But is there any further way of testing the idea against linguistic intuitions, corresponding to the testing of compositionality against cases of substitution? The answer is in fact yes. Corresponding to the functional version of inverse compositionality there is a substitution version:

(IPCS) If the expressions e_1 and e_2 are non-synonymous, and the expression S' results from replacing e_1 by e_2 in S , then S and S' are non-synonymous.

IPCS is in an intuitive sense the inverse of PCS. According to PCS, a substitution that preserves of the meaning of parts also preserves the meaning of the whole. According to IPCS, a substitution that changes the meaning of the parts also changes the meaning of the whole. Moreover, as PCF is equivalent with PCS, under certain conditions to be specified, so IPCF is equivalent to IPCS, under certain conditions to be specified. That this is so, is not as obvious as in the case of PCF and PCS. It will be proved in the next section.

I shall say call a semantics that satisfies the IPCF/IPCS principle *inversely compositional*. A semantics that is both compositional and inversely compositional will be said to be *bidirectionally compositional*, or *bicompositional*. A natural language semantics, or an ordinary formal language semantics, that is bicompositional is also what I shall call *strongly compositional*. I shall return to this notion at the end of the next section.

4. Bidirectionality and strong compositionality: a formal presentation

In the following presentation I shall use the algebraic framework developed by Wilfrid Hodges and, following Hodges, Dag Westerståhl.⁹ This is a simplified version of for-

8. If certain kinds of linguistic constructions aren't semantically compositional, it can still be the case the compositionality offers the correct explanation of why hearers correctly interpret utterances of new sentences that belong to a (big) compositional fragment of a natural language. The compositionality of natural language need not be an all or nothing issue.

mal frameworks in the tradition of Richard Montague's 'Universal Grammar', further developed by Theo Janssen and Hermann Hendriks.¹⁰ For a brief comparison of the two frameworks, see Westerståhl *unpublished*. In both cases, concepts and methods from Universal Algebra are applied for characterizing syntax and semantics.

Let's first define the syntactic algebra:

Definition 1

A language L is a triple (E, A, \circ) where

E is the set of expressions of L

A is the set of *atomic expressions* of L , a subset of E

\circ is the set of syntactic operations, such that each \circ_n is a partial map from E^n to E where n is the arity of \circ_n .

The set \circ is, or corresponds to, the set of syntactic rules of the language L . In the Hodges-Westerståhl framework, there are no basic syntactic categories. Instead, the algebra is partial, in that an n -ary operation \circ_n need not be defined for all members of E^n . One can reconstruct syntactic categories from the domains of the members of \circ , by considering substitutivity classes of expressions.

The set E of L expressions is *generated* from A . That is, E can be inductively defined as the closure of A under the operations in \circ .

Definition 2

With respect to an algebra (B, C, \circ) with the generating set $C \subseteq B$, where $a \in B$,

$C = \{a\}$, $Gen(a, \circ)$, or a is generated from C by means of \circ , iff

a) $a \in B$, $a \in C$, and $a = a$, or

b) $a = \circ_n(a_0, \dots, a_{n-1})$, $Gen(a_i, H_i, K_i)$, for some $H_i, K_i, i < n$, $a_i \in H_i$, and $a_i \in K_i$.

Using ' $[A]$ ' to denote the set of elements generated from A , we thus require that $E = [A]$. Intuitively, A is the set of *simple* expressions of the language. It is natural to require that A be *minimal*, in the sense that there is no proper subset $A' \subset A$ such that $[A'] = [A]$. Also, if both A and \circ are finite, we say that L is a *finitely generated* language. Here we consider only finitely generated languages. Thus, L is finitely generated. The set E is normally infinite, but need not be.

By the term 'complex expression' we shall mean members of $E - A$. Complex expressions may be syntactically ambiguous. That is, a complex expression may be generated

9. Hodges 1998, 2001, Westerståhl *forthcoming, unpublished*.

10. Montague 1970, Janssen 1984, 1997, Hendrix 2001.

in more than one way from the atomic expressions. Typically, a syntactically ambiguous expression may also be semantically ambiguous, in that different meanings are assigned to it corresponding to the different ways it is generated. Therefore, in the algebraic approaches, meanings are assigned to *derivations* of expression rather than to the expressions themselves. The derivations are represented by derivation terms, such as $\langle (e_1, e_2) \rangle$, with $e_1, e_2 \in A$.

In the Hodges-Westerståhl framework the derivation terms of expressions are handled in two different derived algebras, the term algebra $T(L)$ and the grammatical term algebra $GT(L)$. $T(L)$ forms an intermediate step, containing both *variables* and terms that do not, because of the partiality of the operations, correspond to any expressions. $GT(L)$ contains neither. Below a set Var of variables x_0, x_1, x_2 , etc., is introduced, and we shall employ the convention of using signs in boldface to denote the same sign in roman font.

Definition 3

The *term algebra* $T(L)$ of L is a triple (T_L, A, \cdot) , where T_L is a set of terms and \cdot a set of operations. T_L is defined inductively by:

- i) Every expression $e \in A$ is in T_L and is an *atomic term*.
- ii) Every variable $v \in Var$ is in T_L and is an *atomic term*.
- iii) Where \cdot is of arity n and t_0, \dots, t_{n-1} are in T_L , $\sigma(t_0, \dots, t_{n-1})$ is in T_L .

\cdot is the set of operations such that, where \cdot is of arity n and t_0, \dots, t_{n-1} are in T_L , $\cdot(t_0, \dots, t_{n-1}) = \sigma(t_0, \dots, t_{n-1})$.

Thus, the value of \cdot as applied to t_0, \dots, t_{n-1} is the complex term $\sigma(t_0, \dots, t_{n-1})$, and t_0, \dots, t_{n-1} are its *immediate constituents* / *subterms*.

Terms have a unique parsing. Not all the terms of T_L correspond to expressions in E (the term $\sigma(t_0, \dots, t_{n-1})$ itself is well-formed anyway). Only the *grammatical* terms correspond to expressions. The grammatical terms evaluate to expressions. There is a function *val* from GT_L to E inductively defined together with the set GT_L of grammatical terms:

Definition 4

The *grammatical term algebra* $GT(L)$ of L is a triple (GT_L, A, \cdot) , where GT_L is a set of terms and \cdot is as in the definition of $T(L)$, but with operations restricted to GT_L . GT_L is defined inductively together with the function *val* from GT_L to E by:

- i) Every expression $e \in A$ is in GT_L and is an *atomic term*; $val(e) = e$.

- ii) Where f is of arity n , t_0, \dots, t_{n-1} are in GT_L and f is defined for $val(t_0), \dots, val(t_{n-1})$, $f(t_0, \dots, t_{n-1})$ is in GT_L and $val(f(t_0, \dots, t_{n-1})) = (val(t_0), \dots, val(t_{n-1}))$.

val is by requirement *surjective*. That is, every expression e of E is the value of some grammatical term t of GT_L . t is a structural analysis of e . In case e is structurally ambiguous it has more than one structural analysis.

A concept that will be of crucial importance is that of a *free algebra*. According to the more abstract definition of this concept, a free algebra (A, F) in a class K of algebras has a generating set X , such that any function f from X to the domain B of some other member (B, G) of K can be extended into a homomorphism f' from (A, F) to (B, G) (see Grätzer 1968, chapter 4). However, here we are going to employ a less abstract definition:

Definition 5

An algebra (C, D, σ) is *free* iff

- a) $C = [D]$
- b) no member of D is in the range of any σ_i ,
- c) where σ_i is of arity n , and defined for a_0, \dots, a_{n-1} ,
 $(a_0, \dots, a_{n-1}) \rightarrow a_i, i < n$,
- d) where σ_i, σ_j are of arity n and m , and defined for a_0, \dots, a_{n-1} ,
 b_0, \dots, b_{m-1} , respectively, $(a_0, \dots, a_{n-1}) = \sigma_i(b_0, \dots, b_{m-1})$ iff
 $n = m, \sigma_i = \sigma_j$ and $a_i = b_i, i < n$.¹¹

It is clear that term algebras are free. Atomic terms are not values of any operation in the algebra; any value of an operation is distinct from its arguments, and operations give the same values only where they are identical and have the same arguments.

Conversely, if an algebra is free, then the elements are built up as complexes from simple parts, or can be treated as either simple or complexes of this kind. We shall come back to this below.

The next step is to introduce the semantics. In the Montague-Janssen-Hendriks framework, the meaning domain itself constitutes an algebra. It has semantic elements, semantic sorts, semantic operators and operator sorts. In the Hodges-Westerståhl framework, on the other hand, it is only assumed that there is a domain M of meanings. There are no initial assumptions about the nature of the elements of M . No operations are assumed, and no structure of this domain. This set-up is simpler, and it suits the present purposes better.

11. See Montague (1970:225).

The semantic interpretation is effected by a *meaning function* $\mu:GT_L \rightarrow M$, assigning meanings from M to grammatical terms (analysed expressions). In the framework, μ is allowed to be partial, i.e. to not assign a meaning to all grammatical terms. Here, however, in order to avoid unnecessary complications, we shall assume that μ is total. To a meaning function μ there corresponds a synonymy relation \sim_μ on GT_L , defined so that $t \sim_\mu s$ iff $\mu(t) = \mu(s)$. \sim_μ is an equivalence relation, and since μ is total, so is \sim_μ .

μ is said to be *compositional*, if, and only if, μ is a *homomorphism* from GT_L to M . Compositionality can be stated in a function format (**F**) and in a substitution format (**S**), corresponding to PCF and PCS. I shall borrow formulations from Hodges (2001:12).

(**F**) There is a function r such that for any μ -meaningful term $t = \langle t_0, \dots, t_{n-1} \rangle$

$$\mu(t) = r(\langle \cdot, \mu(t_0), \dots, \mu(t_{n-1}) \rangle)$$

With respect to $GT(L)$ and μ we shall pick out one particular function in the set of functions satisfying the **F** equation, viz. the minimal element:

Definition 6

$$= \{r: \text{for any } \mu\text{-meaningful term } t = \langle t_0, \dots, t_{n-1} \rangle, \mu(t) = r(\langle \cdot, \mu(t_0), \dots, \mu(t_{n-1}) \rangle)\}.$$

It is immediate from this definition that itself satisfies the **F** equation, i.e. that for any μ -meaningful term $t = \langle t_0, \dots, t_{n-1} \rangle$,

$$\mu(t) = \langle \cdot, \mu(t_0), \dots, \mu(t_{n-1}) \rangle.$$

Note that $\langle \cdot, \mu(t_0), \dots, \mu(t_{n-1}) \rangle$ is a function from M^n to M . For brevity we shall write it ' $\langle \cdot \rangle$ '. The minimality of guarantees the following:

Fact 1

If $\langle \cdot, \mu(t_0), \dots, \mu(t_{n-1}) \rangle$ is defined, then there is a term $s = \langle \cdot \rangle_i(s_0, \dots, s_{n-1}) \in GT_L$, such that $s_i \sim_\mu t_i$, $i < n$, $\langle \cdot \rangle = \langle \cdot \rangle_i$, and $\mu(s) = \langle \cdot, \mu(t_0), \dots, \mu(t_{n-1}) \rangle$.

Proof: If $\langle \cdot, \mu(t_0), \dots, \mu(t_{n-1}) \rangle$ is defined, but there is no term $s = \langle \cdot \rangle_i(s_0, \dots, s_{n-1}) \in GT_L$, such that $s_i \sim_\mu t_i$, $i < n$, $\langle \cdot \rangle = \langle \cdot \rangle_i$, and $\mu(s) = \langle \cdot, \mu(t_0), \dots, \mu(t_{n-1}) \rangle$, then there is a function ' \cdot ' that satisfies the **F** equation but isn't defined for $\langle \cdot, \mu(t_0), \dots, \mu(t_{n-1}) \rangle$. But this contradicts the definition of \cdot . *QED*

For the next formulation we shall follow Hodges and Westerståhl in using the substitution notation of the type ' $t(t_0, \dots, t_{n-1} | x_0, \dots, x_{n-1})$ ', referring to the term that is formed from the term t of T_L by replacing each variable x_i by the term t_i , $i < n$. Each of the variables is supposed to occur in t only once. Accordingly, the terms t_0, \dots, t_{n-1} need not be

distinct; if one subterm occurs more than once the list will contain repetitions. With the help of this notation we can formulate the substitution version:

(S) If t is a term and $s_0, \dots, s_{n-1}, t_0, \dots, t_{n-1}$ are grammatical terms such that $t(s_0, \dots, s_{n-1} | x_0, \dots, x_{n-1})$ and $t(t_0, \dots, t_{n-1} | x_0, \dots, x_{n-1})$ are both μ -meaningful and for each $i < n$ it holds that

$$s_i \sim_{\mu} t_i$$

then

$$t(s_0, \dots, s_{n-1} | x_0, \dots, x_{n-1}) \sim_{\mu} t(t_0, \dots, t_{n-1} | x_0, \dots, x_{n-1}).$$

The domain of μ , $dom(\mu)$, is said to be *closed under subterms* if every subterm of a μ -meaningful term is also μ -meaningful. It is fairly straightforward to prove the following:

Fact 2 (Hodges 2001)

If $dom(\mu)$ is closed under subterms, then **F** and **S** are equivalent.

F is a close counterpart to the informally stated PCF. **S**, however, differs from PCS partly in that it mentions a range of subterms rather than a single sub-expression to be substituted. This difference is not always negligible. The reason is that there is not in general any guarantee that meaningfulness is preserved under substitution of synonymous expressions. Two terms that can be interchanged in all contexts with preserved meaningfulness (*salva significatione*) are said to belong to the same *semantic category*. Following Hodges (2001:10-11) I shall say that s_0 and t_0 belong to the same μ -category, $s_0 \sim_{\mu} t_0$, if and only if for every term t with the single variable x , $t(s_0 | x)$ is μ -meaningful iff $t(t_0 | x)$ is μ -meaningful. Following a remark in Tarski (1935:215ff), referring to Husserl's notion of a meaning category, Hodges calls such a meaning function μ *Husserlian*:

Definition 7

A meaning function μ is *Husserlian* iff it holds for all terms s and t , that

$$\text{if } s \sim_{\mu} t, \text{ then } s \sim_{\mu} t.$$

If a meaning function μ fails to be Husserlian,¹² then a substitution principle for one subterm at a time is not equivalent with one that concerns an arbitrary number. It may be that replacing one subterm by a synonymous term leads to loss of meaningfulness,

12. When this happens in natural language, it is usually because grammaticality isn't preserved (third-person singular forms of verbs are plausibly held to be synonymous with their first- and second-person counterparts, but cannot replace them without loss of grammaticality). In principle it could happen even within the set of grammatical expressions (in that case the meaning function would not be total).

and then there is no meaning to be preserved by the next substitution. The weaker substitution principle is called ‘1-comp’ by Hodges, and will be called **1-S** here:

- (1-S)** If t is a term and s_0 and t_0 are grammatical terms such that $t(s_0 | x_0)$ and $t(t_0 | x_0)$ are both μ -meaningful and
- $$s_0 \mu t_0$$
- then
- $$t(s_0 | x_0) \mu t(t_0 | x_0)$$
- whenever both these terms are μ -meaningful.

Then the following is straightforward:

Fact 3 (Hodges 2001)

If μ is Husserlian, then **S** and **1-S** are equivalent.

A homomorphic mapping μ from $GT(L)$ to a set of meanings M induces a *semantic algebra* $\mu(GT(L))$, the *homomorphic image* of $GT(L)$ under μ , given the definition of :

Definition 8

S is the algebra $\mu(GT(L)) = (\mu(GT_L), \mu(A), (\cdot))$.

The main interest here is the relation between this algebra and the inverse substitution property, ICPS. The immediate formal counterpart to ICPS concerns substitution of one subterm at the time, corresponding to **1-S**. We let $s \mu t$ mean that μ is defined for s and t and $\mu(s) \mu(t)$.

- (1-IS*)** If t is a term and s_0 and t_0 are grammatical terms such that $t(s_0 | x_0)$ and $t(t_0 | x_0)$ are both μ -meaningful and
- $$s_0 \mu t_0$$
- then
- $$t(s_0 | x_0) \mu t(t_0 | x_0).$$

Corresponding to **1-S** and **1-IS***, we have

- (IS*)** If t is a term and $s_0, \dots, s_{n-1}, t_0, \dots, t_{n-1}$ are grammatical terms such that $t(s_0, \dots, s_{n-1} | x_0, \dots, x_{n-1})$ and $t(t_0, \dots, t_{n-1} | x_0, \dots, x_{n-1})$ are both μ -meaningful and for one or more $i < n$ it holds that
- $$s_i \mu t_i$$
- then
- $$t(s_0, \dots, s_{n-1} | x_0, \dots, x_{n-1}) \mu t(t_0, \dots, t_{n-1} | x_0, \dots, x_{n-1}).$$

IS* is the reversal of **S**. However, **IS*** places a restriction only on terms of the same structure. It allows terms of different structure to be synonymous. We will need a stronger principle. For stating it, two more definitions are needed.

Definition 9

$\dot{\ }_i$ and $\dot{\ }_j$ of arity n are μ -equivalent iff for some n they are both of arity n and for all terms t_0, \dots, t_{n-1} it holds that $\mu(\dot{\ }_i(t_0, \dots, t_{n-1})) = \mu(\dot{\ }_j(t_0, \dots, t_{n-1}))$, or else neither operator is defined for the terms.

$\dot{\ }_i$ and $\dot{\ }_j$ may be μ -equivalent but still different, provided $\dot{\ }_i$ and $\dot{\ }_j$ give values with different surface structure. μ -equivalence is an equivalence relation on $\dot{\ }$.

For the following definition, let ' $s \nabla t$ ' mean that neither of s and t is a proper subterm of the other.

Definition 10

Terms $s, t \in GT_L$ are μ -congruent, $s \mu t$, iff

- a) s or t is atomic, $s \mu t$, and $s \nabla t$, or
- b) for some n , $s = \dot{\ }_i(s_0, \dots, s_{n-1})$, $t = \dot{\ }_j(t_0, \dots, t_{n-1})$, $\dot{\ }_i$ and $\dot{\ }_j$ are μ -equivalent and for $k < n$, $s_k \mu t_k$.

Fact 4

$s \mu t$ implies $s \mu t$.

Proof: By induction on complexity of the left term. The base step is immediate. The induction step proceeds by induction on complexity of the right term. Again the base step is immediate. For the induction step it needs to be shown that if $\dot{\ }_i$ and $\dot{\ }_j$ are μ -equivalent and for $k < n$, $s_k \mu t_k$, then,

$$s \mu t,$$

where $s = \dot{\ }_i(s_0, \dots, s_{n-1})$, $t = \dot{\ }_j(t_0, \dots, t_{n-1})$. Since μ is total, $\mu(s)$ and $\mu(t)$ are defined. By the induction hypothesis we have $s_k \mu t_k$, for $k < n$. Therefore, since μ is compositional,

$$\mu(\dot{\ }_i(s_0, \dots, s_{n-1})) = \mu(\dot{\ }_i(t_0, \dots, t_{n-1}))$$

by **S**. And since $\dot{\ }_i$ and $\dot{\ }_j$ are μ -equivalent, by Definition 9

$$\mu(\dot{\ }_i(t_0, \dots, t_{n-1})) = \mu(\dot{\ }_j(t_0, \dots, t_{n-1}))$$

which is to say that $s \sim_{\mu} t$.

QED

Then we can give the stronger substitution version of inverse compositionality:

(IS) If $s \sim_{\mu} t$, where $s, t \in GT_L$, then $s \sim_{\mu} t$.

The **IS** principle is very close to (the only-if part) of Carnap's definition of synonymy as *intensional isomorphism* (Carnap 1956:56-57). According to Carnap's definition of this notion, two expressions are intensionally isomorphic just if they are atomic and have the same intension, or are complex, have the same structure and their respective simple parts have the same intension. Therefore, it might be useful to be clear about the difference.

First, by means of his notion of intensional isomorphism, Carnap gave a *definition* of synonymy, intended to be so much stronger than sameness of intension so as to support intersubstitutivity *salva veritate* in belief contexts. By contrast, the **IS** principle assumes a prior notion of synonymy. Applied to a natural language, like English, it is either factually true or factually false, given such a prior notion.¹³

Second, **IS** is less restrictive in allowing a simple and a complex term (like 'brother' and 'mail sibling—i.e. ' ('male', 'sibling')—to be synonymous.

Third, **IS** is less restrictive in allowing terms of different syntactic form to be synonymous, given that their syntactic forms are semantically equivalent.

Fourth, **IS** is formulated for grammatical terms rather than for the linguistic expressions corresponding to these terms.

Turning to the relation between **IS** and **IS***, we can note that the former is stronger.¹⁴

Fact 5

IS implies **IS***.

Proof: The contrapositive of **IS*** says that *if*

(*) $t(s_0, \dots, s_{n-1} | x_0, \dots, x_{n-1}) \not\sim_{\mu} t(t_0, \dots, t_{n-1} | x_0, \dots, x_{n-1})$,

then

(**) $s_i \not\sim_{\mu} t_i$, for $i < n$.

13. Of course, data about synonymy are soft, as is illustrated by the discussions in section 7.

14. Note that a version of compositionality (which concerns sufficient conditions for sameness of meaning) that says that the meaning of an expression is a function of the entire structure and the meaning of its *simple* parts is a *weaker* principle than **S**. By contrast, in the case of inverse compositionality (which concerns necessary conditions for sameness of meaning), you get a *stronger* principle by requiring sameness of entire structure and sameness of meaning of simple parts.

This would be a direct consequence of **IS**, if the variables x_0, \dots, x_{n-1} were all immediate constituents of t . But they need not be. Still, since t is complex, $t = \iota_i(u_0, \dots, u_{m-1})$, for some i and some terms u_0, \dots, u_{m-1} . Each variable x_j , for some $j < n$, is a subterm, proper or not, of u_l , for some $l < m$. If $u_l = x_j$, then it is a direct consequence of **IS** that the instance $s_j \mu t_j$ of (***) is true. If u_l contains x_j as a proper subterm, then let $x_{l_0}, \dots, x_{l_{h-1}}$ be the variables among x_0, \dots, x_{n-1} that are subterms of u_l . It is a direct consequence of **IS** that if (*) is true, then

$$(***) \quad u_l(s_{l_0}, \dots, s_{l_{h-1}} \mid x_{l_0}, \dots, x_{l_{h-1}}) \mu u_l(t_{l_0}, \dots, t_{l_{h-1}} \mid x_{l_0}, \dots, x_{l_{h-1}}).$$

Now we apply **IS** again to (***), and repeat the process until x_j is reached, and similarly until all the instances of (***) are derived. *QED*

IS* is not equivalent with **1-IS***, even if μ is Husserlian, for the simple reason that unlike the relation *same as*, the relation *different from* isn't transitive (two substitutions in accordance with **1-IS*** may lead back to the same meaning). The question of the truth of these principles for English will be discussed in section 7.

As was noted above, if we are interested in determining expressions from meanings, the existence of synonyms is a problem. One can avoid this problem technically in two ways. One is to deal with a fragment of the original language that results from filtering out synonyms. The resulting fragment $L' = (E', A', \cdot')$ is such that each equivalence class $\{e: e \mu e_0\}$, where $e_0 \in E'$, contains only one member.

The alternative is to consider instead the *quotient algebra* $Q = GT(L)/\mu = (GT_L/\mu, A/\mu, \cdot/\mu)$ where the elements of GT_L/μ are the equivalence classes $\bar{t} = \{s: s \mu t\}$, $t \in GT_L$, i.e. the synonymy classes, themselves.

Definition 11

The *quotient algebra* Q is the triple $(GT_L/\mu, A/\mu, \cdot/\mu)$, where
 $GT_L/\mu = \{\bar{t}: t \in GT_L\}$,
 $A/\mu = \{\bar{t}: t \in A\}$,
 $\cdot/\mu = \{\bar{\cdot}: \cdot\}$.

Here, where $\bar{\cdot} = ((t_0, \dots, t_{n-1}) = t, \bar{\cdot} / \mu$ is defined so that $\bar{\cdot}(t_0, \dots, t_{n-1}) = \bar{t}$. It is a well-known fact in Universal Algebra that where μ is a homomorphism, so that μ is a congruence relation, then the corresponding quotient algebra is well-defined (cf. Grätzer 1968: 35-36).

To this we add a derived meaning function $\bar{\mu}$ from GT_L/μ to M , defined so that $\bar{\mu}(\bar{t}) = \mu(t)$. Then it holds that

Fact 6

$\bar{\mu}$ is an isomorphism from Q to S .

Proof: Follows by direct application of the Homomorphism Theorem (Grätzer 1968: 57). *QED*

Since $\bar{\mu}$ is an isomorphism, it has an inverse $=\bar{\mu}^{-1}$. $\bar{\mu}^{-1}$ is then a homomorphism from $\mu(GT_L)$ to GT_L/μ , i.e. from meanings to the corresponding synonymy classes. Thus, where $=\bar{\mu}^{-1}$, and $t = \langle t_0, \dots, t_{n-1} \rangle$, it holds that

$$(\mu(t)) = \bar{t}.$$

This fact depends only on the compositionality of μ , and not on **IS**. Whether it is cognitively significant depends on the nature of the synonymy class algebra Q . It is significant if, and only if, the possible syntactic difference between synonymous expressions is restricted in a relevant way. However, **IS** spells out precisely this restriction. If **IS** is true, then any synonymy class \bar{t} is such that all its members are μ -congruent. This suggests how the functional counterpart to **IS** should be formulated.

Requiring that the meaning function μ be 1-1 is too strong, and requiring that the function from synonymy classes to meanings, $\bar{\mu}$, be 1-1 is too weak, since that condition is always met by whenever the quotient algebra is well-defined. A natural middle course would be to take equivalence classes of μ -congruent expressions as the range of the inversely compositional function. That is, with μ° as the inverse function, we would have, for some term s , that

$$\mu^\circ(m) = \bar{s} = \{ t \in GT_L : (\mu(s) = m) \ \& \ (t \sim_\mu s) \}.$$

However, \bar{s} is not in general an equivalence class. Some atomic term might be μ -congruent with two complex terms that are not μ -congruent with each other. If **IS** is true, then the \bar{s} classes *are* equivalence classes, but not otherwise. So, if **IS** is not true, we will have, for some m , t and s , $\mu(t) = m$, $\mu(s) = m$, but $\bar{t} \not\sim \bar{s}$. Then, there is no inverse function μ° .

We can therefore state

(IF) $=\bar{\mu}^{-1}$ is a function from $\mu(GT_L)$ to $\{ \bar{t} : t \in GT_L \} = \{ \bar{t} : t \in GT_L \}$, and there is a function $\bar{\mu}^{-1} : (\cdot) / \mu$, such that, where $t = \langle t_0, \dots, t_{n-1} \rangle$,

$$(\bar{\mu}^{-1}(\mu(t_0), \dots, \mu(t_{n-1}))) = (\bar{\mu}^{-1})(\mu(t_0), \dots, \mu(t_{n-1})).$$

Fact 7

If **F** is true, then **IS** and **IF** are equivalent.

Proof: Since $\bar{\cdot}$ is 1-1, because of Fact 6, it follows that there is a function $\bar{\mu}$ satisfying the equation of **IF**. For in order to produce a counterexample we would need $t_i, t_j \in GT_L$ such that $\bar{\mu}(t_i) \neq \bar{\mu}(t_j)$, but

$$(\bar{\cdot}(\mu(t_0), \dots, \mu(t_i), \dots, \mu(t_{n-1}))) = (\bar{\cdot}(\mu(t_0), \dots, \mu(t_j), \dots, \mu(t_{n-1}))).$$

Since $\bar{\cdot}$ is 1-1, $\bar{\mu}(t_i) \neq \bar{\mu}(t_j)$ iff $\mu(t_i) \neq \mu(t_j)$, and so no counterexample can be produced. What doesn't follow from Fact 6 is that the range of $\bar{\mu}$ is $\{\bar{t} : t \in GT_L\}$. If it is, then $\{\bar{t} : t \in GT_L\} = \{\bar{t} : t \in GT_L\}$, and hence for any s and t that

$$(\#) \quad \text{if } \bar{s} = \bar{t}, \text{ then } \bar{t} = \bar{s},$$

for if (#) doesn't hold for some t_i and t_j , then $\bar{t}_i \neq \bar{t}_j$, even though $\bar{t}_i = \bar{t}_j$, and hence $\bar{t}_i, \bar{t}_j \notin \{\bar{t} : t \in GT_L\}$. Hence, **IS** holds, and that takes care of the right-to-left part. For the left-to-right part, it immediately follows from **IS** that (#) is true, and hence that $\{\bar{t} : t \in GT_L\} = \{\bar{t} : t \in GT_L\}$, which together with the existence of a function $\bar{\mu}$ satisfying the **IF** equation, gives **IF**. *QED*

We can in fact identify the function $\bar{\mu}$ as $\bar{\cdot} / \mu$ such that $(\bar{\cdot}(\bar{\cdot})) = \bar{\cdot}$. This, again, relies on the compositionality of μ . But given this definition of $\bar{\mu}$, it holds that

$$\begin{aligned} (\text{IF}') \quad \bar{\mu}^{-1} & \text{ is a function from } \mu(GT_L) \text{ to } \{\bar{t} : t \in GT_L\} = \{\bar{t} : t \in GT_L\}, \text{ and such} \\ & \text{that, where } t = \bar{\cdot}_i(t_0, \dots, t_{n-1}), \\ & \mu(t) = \mu(\bar{\cdot}_i(t_0, \dots, t_{n-1})) \\ & = (\bar{\cdot}(\mu(t_0), \dots, \mu(t_{n-1}))) \\ & = (\bar{\cdot})(\mu(t_0), \dots, \mu(t_{n-1})) \\ & = \bar{\cdot}(\bar{t}_0, \dots, \bar{t}_{n-1}) = \bar{t}. \end{aligned}$$

The cognitive and epistemic significance of this is that, if you know a meaning m , and you can find $\bar{\mu}(m)$, then you can identify, up to μ -congruence, the expression of m .¹⁵

Before turning to the semantic algebra, we can note the distinction between *bidirectional* and a *strong* compositionality.

Example. Suppose we have the following two algebras:

15. Since μ -congruent sentences can differ only marginally, the task of selecting among them is not of great cognitive significance. For related comments, see section 6.

$$\begin{array}{ll}
P_1=(O_1,A_1, _1), \text{ where} & P_2=(O_2,A_2, _2), \text{ where} \\
O_1=\{a,b,c\} & O_2=\{ _ , _ , _ \} \\
A_1=\{a,b\} & A_2=\{ _ , _ \} \\
_1=\{ _1 \}, \text{ such that} & _2=\{ _2 \}, \text{ such that} \\
_1(a,b)=c, \text{ and } _1(a,c)=b. & _2(_ , _)= _ , \text{ and } _2(_ , _)= _ .
\end{array}$$

Then there is a bijection $_ : P_1 \rightarrow P_2$ such that $_(a)= _ , _(b)= _ ,$ and $_(c)= _ .$

Both $_$ and $_^{-1}$ are homomorphisms (e.g. $_(_1(a,b))= _2(_(a), _(b))$, and $_(_1(a,c))= _2(_(a), _(c))$). Suppose that P_1 is an algebra of expressions, i.e. a language in the present algebraic sense, and P_2 an algebra of meanings. In a slightly extended sense, compared with the definitions above, we can say that $_$ is both compositional and inversely compositional. The $_$ equivalence classes only have one member each, and therefore we can pick these members instead of the singleton sets as the range of the inverse function, i.e. select $_^{-1}$ as the inverse function.

However, P_1 is not a free algebra. b is in the range of $_1$, and so clause b) of Definition 5 isn't met (since this holds for c as well, it doesn't help to exchange $\{a,b\}$ for $\{a,c\}$ as generating set). Because of the isomorphism, the same goes for P_2 .

Definition 12

A mapping μ is *strongly compositional* iff it is both bicompositional and its domain is a free algebra.¹⁶

As the example shows a meaning function μ may be bicompositional, in the extended sense, without being strong. However, the meaning functions that we consider interesting are strong if they are bicompositional, since they are defined for the grammatical terms, which do form a free algebra. For instance, the meaning function $_*$ for the grammatical term algebra associated with P_1 does not satisfy **IS**, since the fact that $_*(b)= _*(_1(a, _1(a,b)))$ violates clause a) of Definition 10.

5. The semantic algebra

We turn now to the nature of the semantic algebra S . We aim to show that if **IS** is true, and the meaning function is Husserlian, then S , or a derivative of S to be defined later, is a free algebra. Definition 5 is the definition of a free algebra.

16. The term 'strong compositionality' has been used before, e.g. in Larson and Segal 1995: 78-79. There strong compositionality is what is here called 'compositionality', whereas 'compositionality' in Segal and Larson is used for the weaker principle that meaning depends on total structure and meanings of simple expressions. Despite this usage, I have adopted the term in want of acceptable alternatives.

Lemma 1

$S=(\mu(GT_L), \mu(A), (\cdot))$, satisfies clause a) of definition 5.

Proof: If S satisfies clause a), then S is generated from $\mu(A)$, i.e. $\mu(GT_L)=[\mu(A)](\cdot)$. To verify this, take an element $m \in \mu(GT_L) - \mu(A)$. By the definition of S , there then exists a term $t \in GT_L$ such that $\mu(t)=m$. Since GT_L is generated from A by means of \cdot , there are terms $t_0, \dots, t_{n-1} \in A$ from which t is generated by means members of \cdot , say $\cdot_0, \dots, \cdot_{k-1}$. Now we can prove by induction over the complexity of subterms of t that $\mu(t)$ is generated from $\mu(t_0), \dots, \mu(t_{n-1})$, by means of $(\cdot_0), \dots, (\cdot_{k-1})$.

The base step is trivial ($\mu(t_0)$ is vacuously generated from $\mu(t_0)$, etc.). Then suppose it holds for complexity levels lower than that of subterm t_j of t , and suppose $t_j = \cdot_1(t_{j_0}, \dots, t_{j_{p-1}})$. By definition of μ ,

$$\mu(t_j) = (\cdot_1, \mu(t_{j_0}), \dots, \mu(t_{j_{p-1}})) = (\cdot_1)(\mu(t_{j_0}), \dots, \mu(t_{j_{p-1}})).$$

Hence, by Definition 2, $\mu(t_j)$ is generated from $\cdot_{i < p} K_{j_i}$, where $\mu(t_{j_i})$ is generated from K_{j_i} , $i < p$. By the induction hypothesis, $K_{j_i} \in \mu(A)$, $i < p$. Hence, $\cdot_{i < p} K_{j_i} \in \mu(A)$. So $\mu(t_j)$ is generated from $\mu(A)$. Hence again, by the induction, this holds for all subterms, including t itself. Hence, m is generated from $\mu(A)$. *QED*

Lemma 2

If **IS** is true and μ is Husserlian, then $S=(\mu(GT_L), \mu(A), (\cdot))$, satisfies clause c) of definition 5.

Proof: If S satisfies clause c), then it holds for any (\cdot) , defined for $m_0, \dots, m_{n-1} \in \mu(GT_L)$, that $(m_0, \dots, m_{n-1}) \in m_i$, $i < n$. We verify this by induction over term complexity. That is, we show by induction that for any term t_i , $i < n$, and terms $t_0, \dots, t_{n-1} \in GT_L$, and any \cdot such that (\cdot) is defined for $\mu(t_0), \dots, \mu(t_{n-1})$,

$$\mu(t_i) \in (\cdot)(\mu(t_0), \dots, \mu(t_{n-1})).$$

This is sufficient. For the base step of the induction, assume that, for $e \in A$, and some (\cdot) defined for $\mu(t_0), \dots, \mu(t_{i-1}), \mu(e), \mu(t_{i+1}), \dots, \mu(t_{n-1})$, that

$$\mu(e) \in (\cdot)(\mu(t_0), \dots, \mu(t_{i-1}), \mu(e), \mu(t_{i+1}), \dots, \mu(t_{n-1})).$$

By Fact 1, it follows that there are terms $s_0, \dots, s_{n-1} \in GT_L$, such that $s_j \in \mu t_j$, $j < n$, in particular $s_i \in \mu e$, and

$$\mu(e) = \mu(\ulcorner (s_0, \dots, s_{n-1}) \urcorner).$$

But since, by assumption, μ is Husserlian, it then also holds that

$$(*) \quad \mu(e) = \mu(\ulcorner (t_0, \dots, t_{i-1}, e, t_{i+1}, \dots, t_{n-1}) \urcorner).$$

But this contradicts **IS**, since by **IS** and (*) $\mu(e)$ and $\mu(\ulcorner (t_0, \dots, t_{i-1}, e, t_{i+1}, \dots, t_{n-1}) \urcorner)$ must be μ -congruent. But by clause a) of Definition 10, they are not, since e is atomic and also a constituent of $\ulcorner (t_0, \dots, t_{i-1}, e, t_{i+1}, \dots, t_{n-1}) \urcorner$.

For the induction step, assume that clause c) is satisfied for all terms of complexity below that of s . Then assume that

$$(**) \quad \mu(s) = \mu(\ulcorner (t_0, \dots, t_{i-1}, s, t_{i+1}, \dots, t_{n-1}) \urcorner).$$

But by **IS** and (**), s and $\ulcorner (t_0, \dots, t_{i-1}, s, t_{i+1}, \dots, t_{n-1}) \urcorner$ are μ -congruent. Since s is a complex term, $s = \ulcorner_1 (s_0, \dots, s_{m-1})$, for some l and m . Since $\ulcorner (s_0, \dots, s_{m-1})$ is μ -congruent with $\ulcorner (t_0, \dots, t_{i-1}, s, t_{i+1}, \dots, t_{n-1}) \urcorner$, it holds by clause b) of Definition 10 and Fact 4 that $m=n$ and that

$$\mu(s_j) = \mu(t_j), j < n, \text{ in particular } \mu(s_i) = \mu(s).$$

But then,

$$\mu(s_i) = \mu(\ulcorner_1 (s_0, \dots, s_{i-1}, s_i, s_{i+1}, \dots, s_{n-1}) \urcorner),$$

and hence

$$\mu(s_i) = (\ulcorner_1)(\mu(s_0), \dots, \mu(s_{i-1}), \mu(s_i), \mu(s_{i+1}), \dots, \mu(s_{n-1})),$$

which contradicts the induction hypothesis. So the assumption is false. *QED*

Lemma 3

If **IS** is true and μ is Husserlian, then $S = (\mu(GT_L), \mu(A), (\ulcorner \urcorner))$, satisfies clause d) of Definition 5.

Proof: By clause d), it holds that where (\ulcorner_1) , (\ulcorner_k) (\ulcorner) are of arity m and n , and defined for $\mu(s_0), \dots, \mu(s_{m-1})$, $\mu(t_0), \dots, \mu(t_{n-1})$, respectively,

$$(\#) \quad (\ulcorner_1)(\mu(s_0), \dots, \mu(s_{m-1})) = (\ulcorner_k)(\mu(t_0), \dots, \mu(t_{n-1})) \text{ iff } \\ m=n, (\ulcorner_1) = (\ulcorner_k) \text{ and } \mu(s_i) = \mu(t_i), i < n.$$

Note that $(\ulcorner_1) = (\ulcorner_k)$ iff for any terms t_0, \dots, t_{n-1} ,

$$(*) \quad (\cdot)_1(\mu(t_0), \dots, \mu(t_{n-1})) = (\cdot)_k(\mu(t_0), \dots, \mu(t_{n-1}))$$

(μ was assumed to be total). But since μ is compositional and Husserlian, this holds iff

$$(**) \quad \mu((\cdot)_1(t_0, \dots, t_{n-1})) = \mu((\cdot)_k(t_0, \dots, t_{n-1}))$$

and again this holds by Definition 9 iff $(\cdot)_1$ and $(\cdot)_k$ are μ -equivalent. Therefore, since μ is compositional and Husserlian, $(\#)$ is equivalent with

$$(***) \quad \mu((\cdot)_1(s_0, \dots, s_{m-1})) = \mu((\cdot)_k(t_0, \dots, t_{n-1})) \text{ iff} \\ m=n, (\cdot)_1 \text{ and } (\cdot)_k \text{ are } \mu\text{-equivalent, and } s_i = \mu t_i, i < n.$$

But if **IS** is true, this follows according to clause b) of definition 10. *QED*

In order for S to satisfy clause b) of Definition 5 it is not enough that $\mu(GT_L)$ is generated from $\mu(A)$. It must also hold that $\mu(A)$ is *minimal*. But this need not be true. The meaning of a syntactic atom need not be a semantic atom. It may be that for some proper subset $A' \subset A$, $[\mu(A')] (\cdot) = [\mu(A)] (\cdot)$. In this case we shall say that $S' = (\mu(GT_L), \mu(A'), (\cdot))$ is a *sub-generating* algebra of S .

Definition 13

$C = (C_1, C_2, (\cdot))$ is a *sub-generating* algebra of $D = (D_1, D_2, (\cdot))$ iff $C_1 = D_1 = [C_2] = [D_2]$, $(\cdot) = (\cdot)$, and $C_2 \subset D_2$.

Definition 14

$A^\circ = A - \{e : e \in A \text{ and } e = \mu t, \text{ for some } t \in GT_L - A\}$.
 $S^\circ = (\mu(GT_L), \mu(A^\circ), (\cdot))$.

That is, we form A° by removing atoms like ‘brother’ that are synonymous with some complex term. This gives us the algebra $S^\circ = (\mu(GT_L), \mu(A^\circ), (\cdot))$.

Lemma 4

If **IS** is true and μ is Husserlian, then $S^\circ = (\mu(GT_L), \mu(A^\circ), (\cdot))$ satisfies clause b) of Definition 5.

Proof: If S° satisfies clause b), then no member of $\mu(A^\circ)$ is in the range of any $(\cdot) \in (\cdot)$. Suppose for *reductio* that $\mu(u) \in \mu(A^\circ)$ for some term $u \in A^\circ$ and that there are $t_0, \dots, t_{n-1} \in GT_L$, and $(\cdot)_k \in (\cdot)$, defined for $\mu(t_0), \dots, \mu(t_{n-1})$, such that

$$(*) \quad \mu(u) = (\cdot)_k(\mu(t_0), \dots, \mu(t_{n-1})).$$

By Fact 1, and since μ is compositional and Husserlian, we have

$$(**) \quad \mu(u) = (\cdot_k)(\mu(t_0), \dots, \mu(t_{n-1})) = \mu(\cdot_k(t_0, \dots, t_{n-1})).$$

Hence, $u = \mu(\cdot_k(t_0, \dots, t_{n-1}))$, and since $\cdot_k(t_0, \dots, t_{n-1}) \in GT_L - A$, it follows that

$$u = \{e : e \in A \text{ and } e = \mu t, \text{ for some } t \in GT_L - A\} = A - A^\circ.$$

But this contradicts the assumption that $\mu(u) = \mu(A^\circ)$.

QED

Lemma 5

If $m \in (GT_L)$ is generated from $\mu(A')$, with $A' \in A$, then there is a term $t \in GT_L$ that is generated from A' such that $\mu(t) = m$.

Proof: By induction over complexity of $\mu(GT_L)$. This is possible, since by Lemma 1, S is generated from $\mu(A)$. The base step, with $m \in \mu(A)$, is immediate from Definition 2. For the induction step, assume that the fact holds for all elements of lower complexity than m , i.e. elements that can be generated from $\mu(A)$ in fewer steps than m .

Then, for some (\cdot_j) of arity n , for some n , and some elements m_0, \dots, m_{n-1} , it holds that

$$m = (\cdot_j)(m_0, \dots, m_{n-1}).$$

But, for some $\cdot_k = (\cdot_k)$, and hence we have

$$m = (\cdot_k, m_0, \dots, m_{n-1}).$$

Then, by Fact 1, there is a term $t = \cdot_j(t_0, \dots, t_{n-1})$, such that $\mu(t_i) = m_i$, $i < n$, $(\cdot_j) = (\cdot_k)$ and $\mu(t) = m$.

Where each m_i , $i < n$, is generated from $\mu(A_i) \in \mu(A)$, it holds that m is generated from $\mu(\cdot_{i < n} \mu(A_i)) = \mu(\cdot_{i < n} A_i)$. By the induction hypothesis it holds for each m_i , $i < n$, that there is a term s_i generated from A_i such that $\mu(s_i) = m_i$. Hence we have that $s_i = \mu t_i$, $i < n$. Since μ is Husserlian and compositional, it follows that

$$\mu(t) = \mu(\cdot_j(t_0, \dots, t_{n-1})) = \mu(\cdot_j(s_0, \dots, s_{n-1})) = m.$$

But then, by Definition 2, $s = \cdot_j(s_0, \dots, s_{n-1})$ is generated from $\cdot_{i < n} A_i$. Since m is generated from $\mu(\cdot_{i < n} A_i)$, the induction step is completed. *QED*

Lemma 6

If **IS** is true and μ is Husserlian, then $S^\circ = (\mu(GT_L), \mu(A^\circ), (\cdot))$ satisfies clause a) of Definition 5.

Proof: It remains to verify that $[\mu(A^\circ)](\cdot) = [\mu(A)](\cdot)$. It suffices to show that for any term $e \in A - A^\circ$, $\mu(e)$ is generated from $\mu(A^\circ)$. Suppose for reductio that $\mu(e) \in \mu(A) - \mu(A^\circ)$ and that $\mu(e)$ is not generated from $\mu(A^\circ)$. Call a term with that property *dangling*.

Since $e \in A - A^\circ$, there is a term $t \in GT_L - A$ and atomic terms $e_0, \dots, e_{n-1} \in A$ such that $e = \mu t$ and $\mu(t)$ is generated from $\{\mu(e_0), \dots, \mu(e_{n-1})\}$. If $e_0, \dots, e_{n-1} \in A^\circ$, then $\mu(e)$ is generated from $\mu(A^\circ)$. Hence, by the *reductio* assumption, some elements in $\{\mu(e_0), \dots, \mu(e_{n-1})\}$, say $\mu(e_i), \dots, \mu(e_{i+j})$, belong to $\mu(A) - \mu(A^\circ)$. But if all of $\mu(e_i), \dots, \mu(e_{i+j})$ are generated from $\mu(A^\circ)$, then so is $\mu(e)$. Hence, at least one element among $\mu(e_i), \dots, \mu(e_{i+j})$ is dangling, as well.

Thus, for each dangling element $\mu(e)$, there is a dangling element $\mu(e')$ contained in the set of elements generating $\mu(e)$. This defines a sequence of dangling elements, $\mu(e), \mu(e'), \mu(e''), \dots$, where each element is generated from a set containing the next. Since generation is transitive, it holds again that for any two elements in $\mu(A) - \mu(A^\circ)$, m_g, m_h , $g < h$, that m_g is generated from a set containing m_h .

But $\mu(A) - \mu(A^\circ)$ is a finite set, say with k elements. Hence, in any segment of $\mu(A) - \mu(A^\circ)$ of length $k+1$, some element $\mu(e_c)$ will occur at least twice. Hence $\mu(e_c)$ is generated from a set $\mu(A_c)$ containing $\mu(e_c)$ itself. By Lemma 5 there is a term $s \in GT_L$ such that

$$(\#) \quad e_c = \mu s$$

and s is generated from A_c . If **IS** is true, it also follows that

$$(\#\#) \quad e_c = \mu s$$

But since $e_c \in A_c$, e_c is a subterm of s . Together with $(\#\#)$ this violates clause a) of Definition 10. Hence, if **IS** is true, there are no dangling elements of $\mu(A)$. Thus, $[\mu(A^\circ)](\cdot) = [\mu(A)](\cdot)$. *QED*

Theorem 1

IS is true and μ is Husserlian, then $S^\circ = (\mu(GT_L), \mu(A^\circ), (\cdot))$ is a free algebra.

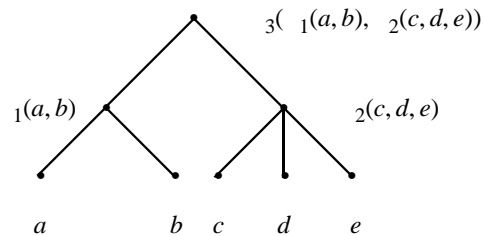
Proof: By lemmas 6 and 4, S° satisfies clauses a) and b) of the definition of a free algebra. By lemmas 2 and 3, $S = (\mu(GT_L), \mu(A), (\cdot))$ satisfies clauses c) and d), but since the properties mentioned there hold for all members of $\mu(GT_L)$ and (\cdot) , and

since these components of S are components of S° as well, S° satisfies clauses c) and d). Hence S° satisfies all defining clauses. *QED*

Next we shall look at the philosophical consequences of Theorem 1.

6. Structured meanings, Frege and the LOT

It is a consequence of Theorem 1 that, if **IS** is true and the meaning function μ is Husserlian, then any element in $\mu(GT_L)$ is generated in a unique way from atomic elements. Because of this, each element is uniquely associated with a finite generating *tree*, where the leaves are all atomic, and where each non-terminal node is formed by an operation in (\cdot) from arguments immediately dominated by it:



This is consistent with the hypothesis that non-atomic meanings are structured, complex entities that are built up from atomic parts. That is, it is consistent with the hypothesis that the tree associated with an element in the semantic algebra represents its constituent structure. A meaning entity is a *part* of the elements that dominate it. Thus, every meaning entity has a unique decomposition into parts. This seems to be what Jerry Fodor has called *reverse compositionality*.¹⁷

We cannot prove (without question-begging premises) that the meanings of complex English expressions are structured entities. What we can (easily) prove is that an algebra consisting of such structured entities is *isomorphic* to S° , and hence that it can serve as the semantic algebra.¹⁸

At this point we can connect back to the end of section 2. There we required of a semantic theory M that it can be combined with a theory of mental representations R , so that the combined theory $M+R$ does explain why linguistic communication succeeds. This condition seems to be met by a theory according to which meanings are structured entities. For such a theory allows a theory R of mental representations according to which it holds that: complex representations isomorphically represent complexes of what

17. Fodor 2000: 371. See also Fodor 1995, 1996. Fodor is, of course, in the first place interested in the decomposition of *representations* of contents.

is represented by their parts. If this is true, and R also, in a corresponding way, relates mental representations to linguistic expressions, then we have an explanation of how a speaker can find an appropriate linguistic expression of a new thought: the speaker's representation of the thought content is systematically associated with a similarly structured representation of the linguistic expression.

To exemplify, suppose a speaker B wants to say *that John is an eye doctor*, and suppose that this is the same as wanting to say *that John is an ophthalmologist*, since these are the same propositional contents. Let's assume an oversimplified syntactic analysis, according to which the sentences

- (1) John is an eye doctor
- (2) John is an ophthalmologist

are analysed as

- (1') σ_1 ('John', '...is an eye doctor')
- (2') σ_1 ('John', '...is an ophthalmologist')

where σ_1 is an operation mapping a singular NP and a singular VP on an S, and assume that the arguments are atomic expressions. Then,

- (3) $val(\sigma_1$ ('John', '...is an eye doctor'))='John is an eye doctor'
- (4) $val(\sigma_1$ ('John', '...is an ophthalmologist'))='John is an ophthalmologist'.

The meaning function μ maps atomic expressions on atomic meanings:

- (5) μ ('John')= m_1
- (6) μ ('...is an eye doctor')= m_2
- (7) μ ('...is an ophthalmologist')= m_2

18. The general idea of structured meanings has been advanced by among others David Lewis (1972) and Max Cresswell (1985). Both distinguish meanings from *intensions*, where intensions are taken to be unstructured entities of possible worlds semantics. For sentences intensions are functions from *indices* to truth values, where indices are possible worlds, points of time and other truth value relevant parameters (Lewis 1972: 173-78). Cresswell, too, has a possible world framework for intensions. He takes an intension of an expression e to be function from entities of appropriate category to sets of possible worlds (Cresswell 1986: 69-70).

Both Lewis and Cresswell think that we do need a more fine-grained concept of meaning, next to that of intension. In Cresswell's case the reason is that we need this for the semantics of indirect discourse. Lewis identifies meanings with trees, where each node consists of an ordered pair of a category and an intension (Lewis 1972: 182-86). Meanings that are parts of more complex meanings are sub-trees. In Cresswell's case structured meanings are ordered tuples of intensions, and such tuples are the referents of *that*-clauses (Cresswell 1985: 101-103).

since the two predicates are assumed to be synonymous, whatever m_1 and m_2 exactly are. There are also functions σ_1 and μ_1 such that

$$\begin{aligned}
 (8) \quad & \mu(\sigma_1(\text{'John'}, \text{'...is an eye doctor'})) \\
 & = (\sigma_1, \mu(\text{'John'}, \mu(\text{'...is an eye doctor'}))) \\
 & = \mu_1(m_1, m_2) = m_3 \\
 & = \text{that John is an eye doctor}
 \end{aligned}$$

$$\begin{aligned}
 (9) \quad & \mu(\sigma_1(\text{'John'}, \text{'...is an ophtalmologist'})) \\
 & = (\sigma_1, \mu(\text{'John'}, \mu(\text{'...is an ophtalmologist'}))) \\
 & = \mu_1(m_1, m_2) = m_3 \\
 & = \text{that John is an eye doctor.}
 \end{aligned}$$

The inverse function σ_1 is then such that

$$(10) \quad \sigma_1(m_1) = \{\text{'John'}\}$$

$$(11) \quad \sigma_1(m_2) = \{\text{'...is an eye doctor'}, \text{'...is an ophtalmologist'}\}$$

and there is a function μ_1 such that

$$\begin{aligned}
 (12) \quad & \mu_1(m_2) = (\sigma_1(m_1), m_2) \\
 & = (\sigma_1(m_1), \sigma_1(m_2)) \\
 & = (\{\text{'John'}\}, \{\text{'...is an eye doctor'}, \text{'...is an ophtalmologist'}\}) \\
 & = \{\sigma_1(\text{'John'}, \text{'...is an eye doctor'}), \sigma_1(\text{'John'}, \text{'...is an ophtalmologist'})\}.
 \end{aligned}$$

B wants to say *that John is an eye doctor*, and since he represents this to himself as a structured content $\mu_1(m_1, m_2)$, he can access his mental lexical entries for m_1 and m_2 . Since m_2 has two entries, corresponding to (6) and (7), the process (more or less) arbitrarily selects one of them, say (7). Applying his internal σ_1 to this selection, and the term-to-expression function *val*, the result is an utterance of (2). The hearer, in turn, can apply his internalized (5), (7) and (9) to get the right interpretation.

The upshot is that a certain kind of meaning function and a certain kind of structure among meanings makes such an explanation possible. Understood this way, I think the present account should be regarded as an exposition of the great passage from Frege:

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 somebody to whom the thought is entirely new. This would be impossible, were we not able to distinguish parts in the thought corresponding to the parts of a sentence, so that the structure of the sentence serves as an image of the structure of the thought (Frege 1923, opening paragraph).

Note that Frege here talks both of the speaker performance and of the hearer performance as something to be accounted for. It is unclear at present whether Frege overstated the case by claiming that (reliably successful) linguistic communication would be *impossible* without this isomorphism between syntax and semantics. This question should be understood as a question whether an *explanation* of the success of linguistic communication is available that does not involve postulating such an isomorphism. It may be that no other explanation is available according to which the internal methods of finding contents and expressions are cognitively symmetric, in the sense of section 1.

This question cannot be fully investigated without considering issues of cognitive architecture. The hypothesis of cognitive architecture that seems to fit in best with the strong compositionality explanation is the *language of thought* hypothesis, LOT. LOT has been suggested and defended by Jerry Fodor in particular in a number of places.¹⁹ According to LOT, mental representations are structured in a language-like syntactic way, with representation of complex entities themselves complex, and built up out of simple representations. Although many questions must be considered, and have been considered, for deciding the issue of cognitive architecture, the present investigation does seem to offer an argument in favour of LOT. The hearer performance in successful linguistic communication involves mapping structured representations (representations of linguistic expressions) on meaning representations, but for that to be possible the meaning representations need not themselves be structured. However, if we take the speaker performance into account, we see that the mapping must be cognitively symmetric, and a cognitively symmetric method is available if the meaning representations are structured as well, i.e. if LOT is true. Any competitor to LOT must match this explanatory advantage.²⁰

19. See, in particular Fodor 1987, appendix, and Fodor 1998, chapter 1.

20. It may be objected, as was suggested to me by Zoltan Szabo, that if the LOT is English (for speakers of English), or rather internal representations of English, then the problem of finding expressions for one's thoughts doesn't arise. Properly speaking, however, it is not that the problem doesn't arise, but that it is solved in the simplest possible way: by the fact the representations of contents and representations of expressions are identical. And whether LOT is English or not, the requirement of strong compositionality of English itself stands.

Another objection made in discussion is that by analogy I should explain our ability to describe what is visually observed in the same way, i.e. by an isomorphism between language and observable reality, which is taken to be absurd. However, what is absurd is only that the structure of observable reality would be *exactly* that of, say, English. Rather, observable reality contains much *more* structure than we take notice of on any single occasion, and probably more than we will ever take notice of. We can e.g. distinguish between a chair and its redness and convey information about this by means of a corresponding English sentence, while disregarding innumerable other parts and properties of the chair. It is not absurd, I think, to assume that a fragment of natural language corresponds to a substructure of experience, or observable reality.

7. Is English inversely compositional?

Unlike the question of whether English is compositional, the question whether English is inversely compositional seems not to have become an object of debate. Here, the discussion can only begin.

At a first glance, English seems to have the property of **1-IS***. Take an arbitrary sentence,

(13) John loves Mary

and exchange some component for a non-synonymous one:

(14) John admires Mary.

The result will clearly be a sentence non-synonymous with the original one. However, already at this level there are possible counterexamples. For it might be claimed that ‘male brother’, as well as ‘male male sibling’ is synonymous with ‘brother’ and ‘male sibling’.²¹ Similarly, ‘red chair’ would be synonymous with ‘red chair’. And in general, iterating an adjectival modifier to a noun head would be seen as not affecting meaning. It is clear that this would refute already the **1-IS*** principle, for ‘sibling’ is certainly not synonymous with ‘male sibling’, but in the context ‘male ...’ they can, on this proposal, be intersubstituted without change of meaning.

Further, the suggestion is incompatible with the present idea of structured meanings. For if $\mu(\text{‘brother’})$ is a proper part of $\mu(\sigma(\text{‘male’}, \text{‘brother’}))$, and they also are identical, then $\mu(\text{‘brother’})$ is a proper part of itself.

But what is the reason for claiming that the two expressions are synonymous rather than merely (necessarily) coextensive? Maybe one reason would be given by linguistic intuitions: the expressions in such pairs simply *seem* to mean the same. However, if intuitions are to be intuition as to how we would interpret utterances of such expressions, then I think they are not synonymous, in fact not even coextensive. The normal interpretation would be to find a contrast that is made precisely by the iteration. In many cases, the iteration has an intensifying effect, as in ‘hot day’ or ‘big big guy’, and even in the case of ‘red red chair’ the effect is intensifying in the sense that the extension must include only chairs that are brightly, intensely or centrally red, as opposed to more peripheral instances of redness. This suggests that in general, the semantics of iteration is a function that takes the extension of the simple head+modifier expression as argument, and gives a central subset of that extension as value, i.e. central or typical with respect to the modifying adjective. Of course, this has no straightforward application in the case of

21. This was suggested by Daniel Nolan and others at the Rutgers semantics workshop in May 2002.

adjectives that don't have a degree, like 'male' or 'even' (as applied to natural numbers), but even here the natural interpretation of such utterances, such as 'male male sibling' or 'even even number' would be to find a contrast along those lines, such as *masculine male sibling*, or *number divisible by four*.

Suppose this is roughly right as a description of speech and interpretation dispositions. It is natural to object this is a merely *pragmatic* phenomenon; it describes how we can use iterations to convey certain information, but it falls strictly outside the semantics of the expressions and constructions used. I think this is a reasonable objection, even though not obviously right. However, if it is conceded that the description is correct by intuitive standards, then intuition does *not* support the idea that the simple and the iterated application of the modifier yields synonymous results. Therefore, appeal to intuition for justifying the synonymy claim is ruled out. To settle the issue, we would have to appeal, if to anything, to questions of simplicity, generality and explanatory power of the semantic theory. If so, it is better to count the expressions as non-synonymous.

Another potential counterexample is provided by conjunctive noun phrases, such as 'Bill and John' and 'John and Bill', where interpretation seems to be insensitive to ordering.²² The idea is that, for some μ , we form $(\text{'John'}, \text{'Bill'}) = \text{'John and Bill'}$ and $(\text{'Bill'}, \text{'John'}) = \text{'Bill and John'}$, which are synonymous, although the terms are not μ -congruent. We can get around this by modifying the theory a little, so that for operators where the order of the arguments doesn't matter semantically, terms count as μ -congruent if the operators are μ -equivalent and the arguments differ at most in order.

But a related problem is worse. For if 'John and Bill and Alfred' can be formed both by $(\text{'John'}, (\text{'Bill'}, \text{'Alfred'}))$ and by $((\text{'John'}, \text{'Bill'}), \text{'Alfred'})$, then the results (the same surface string, under two different analyses) should be counted as synonymous, but the terms are not μ -congruent even under the extended definition just suggested. The immediate subterms are not the same, regardless of order.

Here, I see two options. One is to concede the point, and say that there is a small deviation from inverse compositionality, and from the idea of structured meanings. The deviation is so small that it doesn't seriously impede the communicative task of the speaker, since different ways of representing the same meaning (corresponding to the two different terms above) will be systematically related and contain the same simple parts anyway. On this view, strict inverse compositionality is an ideal from which a semantic theory can depart in some ways without losing much, or anything, in explanatory power.

The other option is to postulate that the μ operator can take arbitrarily many arguments (a multigrade operator), but does not accept other μ -terms in its argument places. This would disallow both $(\text{'John'}, (\text{'Bill'}, \text{'Alfred'}))$ and $((\text{'John'}, \text{'Bill'}), \text{'Alfred'})$,

22. This was suggested by Zoltan Szabo in his comment at the Rutgers semantics workshop.

and leave us with only ('John','Bill','Alfred'). As far as I can see, both options are acceptable.

A third potential counterexample is provided by active-passive transformations. Suppose that

(15) Mary is loved by John

is synonymous with (13), and that they do not have a shared underlying form. Then we have two grammatical terms,

(16) χ_2 ('John', χ_3 ('loves','Mary'))

(17) χ_2 ('Mary', χ_3 ('is loved by','John')).

These terms would correspond two different objects in the semantic algebra, say

(18) $\mu_2(\mu_3(\mu_1('John'), \mu_3(\mu_1('loves'), \mu_1('Mary'))))$

(19) $\mu_2(\mu_1('Mary'), \mu_3(\mu_1('is loved by'), \mu_1('John')))$.

These two terms are clearly different, but on the assumption that (13) and (15) are synonymous they would have the same value, say the proposition *that John loves Mary*, being the same as the proposition *that Mary is loved by John*. So this meaning would decompose in two distinct ways. It could not be identified with any particular structured entity.

Again, there are basically two options for avoiding this conclusion. The first is to declare (13) and (15) non-synonymous but (necessarily) equivalent. I guess this would be preferred for pairs like

(20) John is to the left of Mary

(21) Mary is to the right of John.

since 'left' and 'right' are grammatically unrelated lexical items. In the case of (13) and (15) one might plausibly deny that there are distinct lexical entries for 'loves' and 'is loved by', and therefore an alternative method would be preferred.

For instance, we could claim that the structure of (15) is

(22) $\mu_4(\chi_2(\chi_3('John'), \chi_3('loves','Mary')))$

where μ_4 is the *passive* transformation. We would then add that μ_4 is μ -equivalent with the null-operation, i.e. the operation μ_0 such that for any term t , $\mu_0(t)=t$. Hence, for any term t for which μ_4 is defined, $\mu(\mu_4(t))=\mu(t)$. Thus, μ_4 does not contribute to

semantic structure. On this alternative, both (13) and (15) have structure (18). This seems to be in accordance with what was proposed in transformational grammar (cf. Chomsky 1965:132, Katz and Postal 1964, chapter 3).

Note, though, that this proposal violates clause 1 of Definition 10 (of μ -congruence), since it will mean that a complex term is synonymous with one of its proper constituents (the only one, in this case). However, without harm one can here weaken this restriction to allow transformations, i.e. semantically insignificant operators, and keep the demand for all others.

Another alternative is to distinguish two distinct NP-VT-NP operations, \mathfrak{f}_5 and \mathfrak{f}_6 , giving the terms

(23) $\mathfrak{f}_5(\text{'John'}, \text{'loves'}, \text{'Mary'}) = (13)$

(24) $\mathfrak{f}_6(\text{'John'}, \text{'loves'}, \text{'Mary'}) = (15)$.

\mathfrak{f}_5 and \mathfrak{f}_6 are then assumed to be μ -equivalent. Hence they are mapped on the same semantic function, say \mathfrak{f}_4 , and do not produce different semantic structures. This again would save inverse compositionality, and structured meanings.²³ However, for the same reason as transformations were introduced by Chomsky in the first place—simplifying the grammar—the former alternative is preferable, since we can then stay with a single NP+VP operation.

At a first glance, then, the hypothesis that English is inversely compositional (or at least almost) does seem defensible.²⁴

Department of philosophy
Stockholm University

23. It is interesting to note that Frege discussed the passive transformation, in Frege 1897, pp 60-61, claiming that the thought expressed is the same, despite the difference in form. Moreover, in this passage Frege seems to take two sentences to express different thoughts only if they can differ in truth value, which is difficult to combine with the idea of structured meanings.

24. The paper has been presented at the Rutgers semantics workshop in May 2002, and at the European Congress of Analytic Philosophy, Lund, June 2002. The main ideas had been presented earlier in seminars in Uppsala and Stockholm. I have probably received more helpful comments than I can remember. I owe much to Dag Westerståhl, who took time to go through the paper in detail and suggested several improvements. I am also grateful to Kathrin Glüer-Pagin for helpful suggestions concerning the philosophical parts. I benefited from comments at the Rutgers workshop, especially from Zoltan Szabo, Daniel Nolan, and Michael Glanzberg. Finally, comments from an anonymous referee of JPL led to a number of improvements.

The research has been supported by funding from the Swedish interdisciplinary research project *Meaning and Interpretation*, led by Dag Prawitz. This project in turn is funded by the Tercentenary Foundation of the Swedish National Bank. I have also received support from Tercentenary Foundation of the Swedish National Bank for the project *Meaning, Communication, Explanation*.

References

- Carnap, R, 1956, *Meaning and Necessity*, second edition, The University of Chicago Press, Chicago.
- Chomsky, N, 1965, *Aspects of the Theory of Syntax*, MIT Press, Cambridge, Mass.
- Cresswell, M, 1986, *Structured Meanings*, MIT Press, Cambridge, Mass.
- Fodor, J, 1987, *Psychosemantics*, The MIT Press, Cambridge, Mass.
- Fodor, J, 1996, 'Connectionism and systematicity (continues): why Smolensky's solution *still* doesn't work', *Cognition* 62: 109-19. Reprinted in Fodor 1998b. Page references to the reprint.
- Fodor, J, 1995, 'Review of Christopher Peacock's *A Study of Concepts*', London Review of Books, April 20. Reprinted in Reprinted in Fodor 1998b. Page references to the reprint.
- Fodor, J, 1998, *Concepts. Where Cognitive Science Went Wrong*, Oxford University Press, Oxford.
- Fodor, J, 1998b, *In Critical condition*, The MIT Press, Cambridge, Mass.
- Fodor, J, 2000, 'Reply to critics', *Mind & Language* 15: 350-74.
- Frege, G, 1897, 'Logik', in Frege, *Schriften zur Logik und Sprachphilosophie. Aus dem Nachlass*, Felix Meiner Verlag, Hamburg, 1978.
- Frege, G, 1923, 'Compound thoughts' ('Gedankengefüge'), *Beträge zur Philosophie des Deutschen Idealismus*, 36-51. Reprinted in Frege, *Logische Untersuchungen*, Vandenhoeck & Ruprecht, Göttingen 1976. Translation by R. Stoothoff published in *Mind* 72: 1-17, 1963.
- Gödel, K, 1931, 'On formally undecidable propositions of Principia Mathematica and related systems I', in J van Heijenoort (ed), *Frege and Gödel. Two Fundamental Texts in Mathematical Logic*, Harvard University Press, Cambridge, Mass. 1970. Originally published as 'Über formal unentscheidbare Sätze der Principia mathematica und verwandter System I', *Monatshefte für Mathematik und Physik* 38: 173-98.
- Grätzer, G, 1968, *Universal Algebra*, D.Van Nostrand Company, Inc. Princeton, NJ.
- Hendrix, H, 'Compositionality and model-theoretic interpretation', *Journal of Logic, Language and Information* 10: 29-48.
- Hodges, W, 1998, 'Compositionality is not the problem', *Logic and Logical Philosophy* 6: 7-33.
- Hodges, W, 2001, 'Formal features of compositionality', *Journal of Logic, Language and Information* 10: 7-28.
- Janssen, T, 1984, *Foundations and Applications of Montague Grammar*, Centre for Mathematics and Computer Science, Amsterdam 1984.
- Katz, J and Postal, P M, 1964, *An Integrated Theory of Linguistic Descriptions*, MIT Press, Cambridge, Mass.
- Janssen, T, 1997, 'Compositionality', in van Benthem, J. and ter Meulen, A. (eds), *Handbook of Logic and Language*, Elsevier, Amsterdam 1997.

- Larson, R and Segal, G, 1995, *Knowledge of Meaning. An introduction to Semantic Theory*. MIT Press, Cambridge, Mass.
- Lewis, D, 1972, 'General semantics', in D Davidson and G Harman (eds), *Semantics of Natural Language*, Reidel, Dordrecht.
- Montague, R, 1970, 'Universal grammar', *Theoria* 36: 373-98. Reprinted in R Thomason (ed), *Formal Philosophy. Selected Papers of Richard Montague*. Yale University Press, New Haven. Page references to the reprint.
- Pagin, P, *forthcoming*, 'Schiffer on communication', forthcoming in *Facta Philosophia*.
- Schiffer, S, 1987, *Remnants of Meaning*, MIT Press, Cambridge, Mass.
- Schiffer, S, 1991, 'Does mentalese have a compositional semantics?', in Barry Loewer and Georges Rey (eds), *Meaning in Mind; Fodor and His Critics*, Blackwell, Oxford.
- Tarski, A, 1935, *Der Wahrheitsbegriff in den formalisierten Sprachen*, *Studia Philosophica*, Leopoli. Translated as 'The concept of truth in formalized languages', in Tarski, *Logic, Semantics, Metamathematics*, J. Corcoran (ed), Hackett Publishing, Indianapolis, 1983. Page references to the 1983 edition.
- Westerståhl, D, *forthcoming*, 'On the compositionality of idioms: an abstract approach', forthcoming in D. Barker-Plummer, D. Beaver, J. van Benthem, P. Scotto di Luzio (eds), *Proceedings of LLC8*, CSLI Publications, Stanford.
- Westerståhl, D, *unpublished*, 'On extensions of compositional semantics: variations on a result by Hodges', draft.