On the Compositionality of Idioms
An Abstract Approach

Dag Westerståhl

A standard view is that idioms present problems for compositionality. The question of compositionality, however, should be posed for a semantics, not for individual phrases. The paper focuses on the *idiom extension problem*: If in a given language a certain phrase acquires the status of an idiom, how can the syntax and semantics be extended to accommodate the idiom, while preserving desirable properties such as compositionality? Various ways to achieve such extensions are discussed within an abstract algebraic framework due to Wilfrid Hodges.

1 Introduction

It is a fairly common assumption that the occurrence of idioms in natural languages means trouble for the principle of compositionality. The meaning of a complex idiom seems typically *not* to be determined by the meaning of its parts and the way they are composed. If you know English, and in particular the meaning of *kick* and *the bucket*, but are unfamiliar with the idiom, there is no way you can compute the (idiomatic) meaning of *kick the bucket*. Like lexical items, idioms have to be learned one by one. Yet they appear to have syntactic structure, so compositionality is in trouble. Or is it?

In this paper I first make a methodological point: the common assumption as expressed above is misleadingly put. The real issue is

*This paper is an extended version of Westerståhl (1999): proofs are included, results are sharpened (e.g., Proposition 7.2 and Corollary 10.2), elucidations have been added, and a proposal from Nunberg et al. (1994) of how to analyze certain idioms has been accommodated (in section 10). I have benefitted from comments by a number of colleagues on various occasions where this material was presented. In particular, I would like to thank David Beaver for his comments after the *LLC8* conference at Stanford, which inspired section 10, and two anonymous referees for very helpful remarks and questions. Work on the paper was partially supported by the Bank of Sweden Tercentenary Foundation in connection with the project *Meaning and Interpretation*.

Submitted for
Proceedings of LLC8
D. Barker-Plummer, D. Beaver, J. van Benthem, P. Scotto di Luzio (eds.)

about what I call the *idiom extension problem*: Can we in a given language accommodate new idioms while preserving compositionality? This is a fairly precise question, and it can be answered. The main part of the paper uses a handy algebraic framework (due to Wilfrid Hodges) to formulate various ways to think about idioms, and to see how the idiom extension problem can be solved in each case.

There is no way I could do justice to the vast linguistic literature about idioms. Instead I will take one paper, Nunberg et al. (1994), as background, because it surveys a fair amount of that literature, and furthermore, it propounds a general view of idioms that seems reasonably widespread, at least in its broad outlines. More precisely, Nunberg et al. exemplify with ample quotations from the literature the common assumption about idioms and compositionality I mentioned above, and for *some* idioms, such as *kick the bucket*, they too subscribe to that view. On the other hand, for many other idioms, of which we can take *pull strings* as a prototype, they claim that compositionality does hold, provided one realizes that the parts or 'chunks' of such idioms also have idiomatic meanings.

2 A Methodological Point

My simple and presumably obvious methodological point is this: *While it makes good sense to ask if a semantics is compositional or not, it makes no sense to ask the same question about a particular phrase.*

The idea that a particular idiom is non-compositional might arise as follows. The meaning of *kick the dog*, say $\mu(\textit{kick the dog})$, is a function, say r, of the meanings of *kick* and *the dog*:

(1.1) $\mu(\textit{kick the dog}) = r(\mu(\textit{kick}), \mu(\textit{the dog}))$.

But, for the idiomatic reading,

(1.2) $\mu(\textit{kick the bucket}) \neq r(\mu(\textit{kick}), \mu(\textit{the bucket}))$.

While this is true, it says nothing about the compositionality principle. That principle only asks that *some* function determine the meaning, not that this function be r. Simply redefining r for the idiomatic case will do the trick.

Perhaps it is retorted that a counterexample to the claim of compositionality for a particular semantics can nevertheless be obtained in the following way: consider the relation $\equiv_\mu$ of synonymy: $p \equiv_\mu q$ iff $\mu(p) = \mu(q)$. Assume that *bucket* $\equiv_\mu$ *pail*. Then

(1.3) *lift the bucket* $\equiv_\mu$ *lift the pail*, as compositionality would have it, but

(1.4) *kick the bucket* $\not\equiv_\mu$ *kick the pail.*

contradicting functionality.

However, this only reveals another trivial point. Meanings cannot be assigned to surface manifestations, because then *kick the bucket* would be *ambiguous*. But ambiguity is not our problem here. Standard formulations of the compositionality principle presuppose single-valuedness of meaning. And indeed there is no disagreement that, at the level where meaning is assigned, the idiomatic version of *kick the bucket* must be distinct from the ordinary version, if only by the presence of an idiomatic 'marker'.

But then, the 'modes of composition' in (1.1) and (1.2) are distinct, so there is no reason to expect the same function to operate in both cases. Similarly, on both sides of the synonymy sign in (1.3), and on the right hand side in (1.4), we have one 'mode of composition', but on the left hand side in (1.4) we have another, so no violation of compositionality occurs.

I do not claim to have found the above 'arguments' in the literature; indeed, once explicitly formulated they would hardly convince anyone. Still, the standard version of the compositionality priciple,

(C) The meaning of a complex expression is determined by the meanings of its parts and the mode of composition,

where 'is determined by' is taken as 'is a function of', has so often been taken to be in *obvious* conflict with the occurrence of idioms.[1] How can that be?

Of course, one may very well claim that there is a *familiar* way, or that there are familiar ways, to compose a transitive verb meaning with an NP meaning to form a VP meaning, and that none of those ways applied to the meanings of *kick* and *the bucket* gives the correct result. But there is no immediate road from this obvious fact to the negation of (C).

Clearly some sort of familiarity with the relevant function or rule is required to explain how we can *figure out*, or *know*, or *understand* the meaning of a complex phrase from the meanings of its parts. And it seems to be characteristic of many idioms that we cannot 'figure out' their meanings at all. One might want to express this by saying that

[1] Note that extra requirements of computability of that function make no difference to the present issue. One idiomatic exception (hence any finite number) can always be taken care of, it seems.

they are not compositional. But, to repeat, that would be misleading: there is no direct link from this observation to (C).

Likewise, when Nunberg et al. argue (I think) that for idioms like *pull strings* we actually *can* use the familiar meaning function, given that in the idiom, *pull* and *string* have non-standard meanings, this is an interesting claim, but it is still misleading to call those idioms 'compositional' and contrast them with 'non-compositional' ones. The question is rather: Are there any obstacles to a reasonable compositional semantics for a language containing both kinds of idioms?

3 Basic Issues about Idioms and Compositionality

Assume again that *bucket* $\equiv_\mu$ *pail*, and suppose we have agreed on a particular 'mode of composition', or rule, or marking, that is at work in the idiomatic version of *kick the bucket*. Can we then avoid, except by ad hoc stipulation, that the same 'mode of composition' is applied to yield also an idiomatic version of *kick the pail*? If not, then, by compositionality, the latter has to mean *die* as well.

This is an instance of *the overgeneration problem*. Note the difference from the alleged argument above. There we (falsely) claimed to have a counter-instance to compositionality by ignoring that two distinct 'modes of composition' were involved. Here we acknowledge the need for an 'idiomatic mode of composition', and ask if our analysis, together with compositionality, forces us to countenance idiomatic expressions that do not exist.

The overgeneration problem has been widely discussed in the literature. Other noted issues concern the fact that certain syntactic operations apply to some idioms but not to others. For example,

(1.5) The bucket was kicked by John yesterday

cannot mean that John died yesterday: this idiom does not allow passivization. Similarly for anaphoric reference:

(1.6) Andrew kicked the bucket in June last year, and a month later, Jane kicked it too.

Here too, the idiomatic meaning is hard or impossible to get. But with the literal reading (1.5) and (1.6) are both fine (if a bit odd). For the idiom *pull strings*, on the other hand, both of these operations are all right:

(1.7) Strings were pulled to secure Henry his position.

(1.8) Kim's family pulled some strings on her behalf, but they weren't enough to get her the job. (Nunberg et al., 1994, 502)

One suggestion here is that *kick the bucket* is somehow an *atomic* expression, in contrast with *pull strings*. A different idea is that the former idiom stands for a one-place predicate, the latter for a two-place one, so it is natural that passivization and anaphoric reference to the object position apply precisely in the latter case. This kind of semantic explanation may (on some accounts) be applied regardless of whether an idiom is seen as having structure or not. If it does have structure, there are still various ways to think about it. One natural idea is that (some) idioms have syntactic but not semantic structure, as it were. An alternative is that they have both. Then (some) idioms are composed of parts (some of) which themselves have idiomatic meanings. All of these suggestions will be considered in this paper.

4 The Idiom Extension Problem

It is useful, I think, to formulate of the problem of the compositionality of idioms in the following general terms. Let a language with a compositional semantics be given. Suppose a complex expression in that language acquires, for reasons we need not go into, an idiomatic meaning. How can the given syntax and semantics be *extended* in a way which is natural, accounts for (some of) the behavior of the new idiom, and preserves compositionality and perhaps other desirable properties too?

In the rest of this paper I will consider the form this problem takes under various ideas about how to analyze idioms, within an abstract algebraic framework introduced in Hodges (2000).[2]

The main point of looking at idioms from such an abstract perspective is generality. Essentially, the framework assumes nothing more

[2] Hodges considers two kinds of extension problems related to compositionality. His main focus is on how to extend a partial semantics, i.e., one giving meaning to some but not all kinds of expressions of the language, to a total compositional one. More precisely, it is shown under which conditions there exists a *fregean* extension of the partial semantics μ. The latter notion includes, besides compositionality and the husserl property (section 5 below), a requirement that, roughly, the extended semantics make no further distinctions than those warranted by μ. This requirement, which seems related to (one interpretation of) Frege's so-called Context Principle, will not be used here.

The second extension problem concerns adding new expressions to the language. The idiom extension problem can be seen as an instance of this. A difference between Hodges's approach and mine is that he treats total semantics as the norm, whereas I make non-trivial use of the familiar idea that an expression may be syntactically derivable but still meaningless.

than that there are rules which generate, from atoms, complex terms ('analysis trees'), to which meanings can be assigned. This is enough to state and solve several versions of the idiom extension problem. And that should be enough, I hope, to demonstrate that there are no obstacles in principle to include idioms in a compositional semantics.

I am not claiming, though, that these results necessarily show *how* idioms are best dealt with compositionally. A particular linguistic theory will have more tools at its disposal than the ones available in the abstract algebraic framework, and may therefore have better ways to account for the occurrence and distribution of idioms. In fact, it is an interesting question how the constructions given here relate to the particulars of theories that do have an account of idioms, such as HPSG (cf. Sag and Wasow 1999) which implements ideas about idioms from Nunberg et al. (1994). I will make a few remarks in this connection, but details have to be left for another occasion. But if it can be shown already in our austere framework that a certain idiom extension *exists*, then certainly it exists in more concrete and detailed theories as well. To establish that general existence claim is the aim of this paper, and hence to dispel the impression, still widespread, that the occurrence of idioms is incompatible with compositionality.

5 An Algebraic Framework

The following definitions are from Hodges (2000), with a few twists that will be noted. The framework is a much simplified version of the term algebra account of syntax and semantics first introduced by Montague and developed in, for example, Janssen (1997).

5.1 Definition.

- A *grammar*

  $$(E, A, \underline{\alpha})_{\alpha \in \Sigma}$$

  consists of a set E of *expressions*, a set $A \subseteq E$ of *atomic* expressions, and for each function symbol $\alpha \in \Sigma$ a corresponding *syntactic rule*: a partial map $\underline{\alpha}$ from E^n to E, for some n.[3]

- Let Var be a set of variables, not belonging to E. The set $T(E)$ of *terms* is defined as follows.

[3] I use 'E' in what follows ambiguously for the grammar and for its set of expressions. Hodges does not need to distinguish explicitly between a symbol α and the corresponding function $\underline{\alpha}$; the reason for doing so here will become apparent in section 9.

– $Var \cup E \subseteq T(E)$
– If $t_1, \ldots, t_n \in T(E)$ and $\alpha \in \Sigma$ is n-ary, then $\alpha(t_1, \ldots, t_n) \in T(E)$.

The terms in $T(E)$, which may contain (meta)variables, are mainly introduced to handle substitution conveniently (cf. below). We are really interested in a subset of $T(E)$, namely, those terms which record derivations of (wellformed) expressions by means of the syntactic rules:

- The set $GT(E)$ of *grammatical terms* and the function $val : GT(E) \longrightarrow E$ are given by:
  – $a \in A$ is an *atomic* grammatical term, and $val(a) = a$.
  – Suppose $\alpha \in \Sigma$ is an n-ary function symbol, and $p_1, \ldots, p_n \in GT(E)$ with $val(p_i) = e_i$. If $\underline{\alpha}(e_1, \ldots, e_n)$ is defined, say $\underline{\alpha}(e_1, \ldots, e_n) = e$, the term $\alpha(p_1, \ldots, p_n)$ is in $GT(E)$, and $val(\alpha(p_1, \ldots, p_n)) = e$.

For example, if the term

$$p = \alpha(a, \beta(b, c)),$$

with $val(p) = e$ is grammatical, this represents the fact that the expression e can be derived by first applying rule $\underline{\beta}$ to the atoms b and c, and then applying rule $\underline{\alpha}$ to a and the expression $e' = val(\beta(b, c))$. That $\beta(b, c)$ is grammatical means precisely that rule $\underline{\beta}$ applies to the arguments b and c, i.e., that $\underline{\beta}(b, c) = e'$ is defined. Likewise, p is grammatical because $\underline{\alpha}(a, e') = e$ is defined.

It is helpful to think of expressions as 'surface like' strings, and grammatical terms as analysis trees. *val* returns the string corresponding to each tree. A term or tree describes which rules have been applied and in which order, i.e., it describes a *derivation* of an expression. If two distinct terms correspond to the same expression ($p \neq q$ but $val(p) = val(q)$) we have a structural ambiguity at the expression level. Thus, meanings should be assigned to grammatical terms, not to expressions. In fact, in the present set-up a semantics amounts to nothing more than such an assignment:

5.2 Definition.

A *semantics* for E is a function μ whose domain is a subset of $GT(E)$. $p \in GT(E)$ is *μ-meaningful* if $p \in dom(\mu)$, and p and q are *μ-synonymous*, $p \equiv_\mu q$, if $\mu(p) = \mu(q)$. $\equiv_\mu$ is an equivalence relation on $dom(\mu)$. Two semantics μ and ν for E are *equivalent* if $\equiv_\mu$ equals $\equiv_\nu$.

It should be noted how little these notions presuppose about syntax and semantics. Virtually nothing is assumed about meanings, except that they are assigned to grammatical expressions. If $p \in GT(E) - dom(\mu)$, then p is grammatical though meaningless (relative to μ); such expressions (terms) are familiar from some conceptions of syntax and semantics.[4]

Similarly, a grammar in our sense may represent many different accounts of syntax. We assume that a level of 'expressions', some of which are atomic, can be isolated, and that the rules of syntax can be described by means of (partial) functions from expressions to expressions. But there is no assumption that these functions provide the 'intension' of the rules, or the generalities about a language these are meant to capture. The rules might concatenate surface strings, or fuse surface strings with appended syntactic/semantic information, or they might perform various destructive operation on their arguments. Only their extension as functions matters here.

So the algebraic framework we are using embodies practically no claims about the correct form of syntax or semantics. Rather, it is a *meta-framework* in which various syntactic/semantic theories can be represented, even though most of what is specific about such a theory gets lost in the representation. However, as Hodges observed, what little remains is enough to say some substantial things about compositionality. We shall see that it is also enough to state some facts about idioms.

First, let us fix the relevant notion of compositionality.

5.3 Definition. μ is *compositional* if $dom(\mu)$ is closed under subterms and for each $\alpha \in \Sigma$ there is a function r_α such that, whenever $\alpha(p_1, \ldots, p_n)$ is μ-meaningful,

$$\mu(\alpha(p_1, \ldots, p_n)) = r_\alpha(\mu(p_1), \ldots, \mu(p_1)).$$

If s, p are terms in $T(E)$, and x is a variable, let

$$s(p|x)$$

be the result of replacing all occurrences of x in s by p.

5.4 Fact [Hodges 2000]. *If μ is compositional then, whenever $s \in T(E)$ and $p, q \in GT(E)$ are such that $p \equiv_\mu q$ and $s(p|x)$ and $s(q|x)$ are both μ-meaningful, $s(p|x) \equiv_\mu s(q|x)$.*

[4] Cf. *Colorless green ideas sleep furiously.*

In fact, as Hodges shows, if $dom(\mu)$ is closed under subterms and μ is husserlian (see below), this condition is *equivalent* to compositionality.

Part of the simplicity of the present concept of grammar comes from the fact that we did not need syntactic categories at the outset. Instead, we exploited the partiality of the functions $\underline{\alpha}$ for $\alpha \in \Sigma$; $\underline{\alpha}$ can be undefined for arguments of the 'wrong' sort. However, such categories can be reintroduced as follows: two terms have the same syntactic (semantic) category if they can be substituted everywhere with preserved grammaticality (meaningfulness). Suppose $p, q \in GT(E)$:

5.5 Definition.

- $p \sim_E q \iff \forall s \in T(E)(s(p|x) \in GT(E) \Leftrightarrow s(q|x) \in GT(E))$
- $p \sim_\mu q \iff \forall s \in T(E)(s(p|x) \in dom(\mu) \Leftrightarrow s(q|x) \in dom(\mu))$.

  $\sim_E$ and $\sim_\mu$ are equivalence relations on $GT(E)$. The syntactic (semantic) *category* of $p \in GT(E)$ is its equivalence class $[p]_E$ ($[p]_\mu$). Cat_E is the set of syntactic categories of E.
- E is *categorial* if, for each $\alpha \in \Sigma$,

  $\alpha(p_1, \ldots, p_n), \alpha(p'_1, \ldots, p'_n) \in GT(E) \Rightarrow p_i \sim_E p'_i,\ 1 \le i \le n.$

Normally one would assume that semantics refines syntax, i.e., that $p \sim_\mu q$ implies $p \sim_E q$ (at least when p and q are μ-meaningful), but I shall not need this assumption here.

The husserl property says that synonymous expressions have the same category. More precisely:

5.6 Definition. μ is (a) *husserlian*, (b) *weakly husserlian* if, correspondingly,

(a) For all $p, q \in GT(E)$, $p \equiv_\mu q$ implies $p \sim_\mu q$.

(b) For all $p, q \in GT(E)$, $p \equiv_\mu q$ implies $p \sim_E q$.

The husserl property turns out to be very handy: it simplifies the definition of compositionality, and it plays an essential role in Hodges's Extension Theorem. In this paper, the weak husserl property, together with the assumption that E is categorial, will be needed to make a certain kind of idiomatic extension work (section 8). Both assumptions appear quite modest. But to verify that they hold one has to check the details of a particular grammar. It is certainly possible to construct grammars for which they fail. For example, if one is not careful, a term like

$$\alpha(\gamma(\mathit{the}, \mathit{men}), \mathit{laugh})$$

could be meaningful, *laugh* synonymous with *laughs*, but

$$\alpha(\gamma(\mathit{the}, \mathit{men}), \mathit{laughs})$$

ungrammatical (i.e., belonging to $T(E) - GT(E)$), violating the weak husserl property. There are many ways in which this particular problem could be avoided, but I will not go into details here.

Also, certain uses of subcategories may make a grammar non-categorial in our sense.[5] This too can be avoided, but one might instead want to make a revision of the latter notion — such a revision seems entirely feasible, but will not be undertaken in this paper.

6 Atomic Extensions

Now we can get to work on the idiom extension problem. Fix a grammar E, a semantics μ for E, and a (complex) term $q_0 = \alpha_0(q_{01}, \dots q_{0k}) \in dom(\mu)$ with surface form

$$e_0 = val(q_0).$$

Suppose that e_0 acquires the status of an idiom, which is to have the meaning m_0. How can E and μ be extended?

We noted that one idea (also hinted at in Hodges 2000) is that some idioms are new atoms, looking like familiar expressions on the surface but in fact without syntactic structure. This is easily modeled in the present framework: assume $e_0 \in E - A$, and consider

$$E^a = (E, A \cup \{e_0\}, \underline{\alpha})_{\alpha \in \Sigma}.$$

Thus, the expressions are the same, and so are the syntactic rules. When a rule encounters the argument e_0, it must (being a function from expressions) treat the case when this expression comes from the atom (i.e., the idiom) and the case when it comes from the complex term q_0 in exactly the same way. So no new rules are needed. The only things that happened is that one expression has been promoted to an

[5]Suppose a category C has words of distinct subcategories C1 and C2, so that there are a of category C1 and b of category C2 with $a \not\sim_E b$. Then we cannot have the same syntactic rule $\underline{\alpha}$ applying to both a and b, if E is categorial.

atom, and that is enough to disambiguate between the idiomatic and the literal reading of e_0.

For example, we could have

$$e_0 = \textit{kick-the-bucket} = val(q_0) = val(\alpha_0(kick, \gamma(the, bucket))).$$

If $\underline{\alpha}$ is a rule which takes an NP and a VP in the infinitive and produces a sentence in the past tense, *John kicked the bucket* can now be derived in two ways, as witnessed by the terms

$$\alpha(John, \alpha_0(kick, \gamma(the, bucket))) \text{ and } \alpha(John, e_0).$$

These terms have the same value (surface form), and they are both grammatical because the respective arguments of α also have the same value.[6]

We call E^a an *atomic extension* of E. $T(E^a)$, $GT(E^a)$, and val^a are uniquely determined by E^a. We have $T(E^a) = T(E)$ and

$$GT(E) = \{p \in GT(E^a) : e_0 \text{ does not occur in } p\} \tag{1.9}$$

q_0 is the literal version of the expression e_0. More generally, we obtain the literal version of any grammatical term p of E^a by replacing e_0 with q_0:

If $p = s(e_0|x)$ where $s \in T(E)$, let $p^{\mathrm{lit}} = s(q_0|x)$.

This is well-defined since s is uniquely determined up to the variable x. Observe that

$$e_0^{\mathrm{lit}} = q_0$$

(take $s = x$), and that if $p \in GT(E)$ then

$$p^{\mathrm{lit}} = p$$

(take $s = p$). The next lemma provides more information about how E^a is related to E.

6.1 Lemma.

(a) *If $p, q \in GT(E^a)$, then $p \sim_{E^a} q \Leftrightarrow p^{lit} \sim_E q^{lit}$.*

(b) *If $p \in GT(E^a)$, then $p \sim_{E^a} p^{lit}$ and $val^a(p) = val(p^{lit})$.*

[6] Note that treating *kick the bucket* as an atom in no way forces us to treat (implausibly) *kicked the bucket*, say, as a *distinct* atom.

(c) *If E is categorial, so is E^a.*

(d) *The map $'$ defined by $([p]_E)' = [p]_{E^a}$ is a bijection from Cat_E to Cat_{E^a}.*

Proof. First, note that by the definition of grammatical terms in E^a and the fact that $e_0 = val(q_0)$, it is practically immediate that

$$e_0 \sim_{E^a} q_0. \tag{1.10}$$

From (1.10) and (1.9) we obtain that whenever s is an e_0-free term,

$$s(e_0|x) \in GT(E^a) \Leftrightarrow s(q_0|x) \in GT(E), \tag{1.11}$$

and it is also clear that $val(s(q_0|x)) = val^a(s(e_0|x))$. From (1.11) it follows more generally that when $p \in GT(E^a)$ and s is e_0-free,

$$s(p|x) \in GT(E^a) \Leftrightarrow s(p^{\mathrm{lit}}|x) \in GT(E), \tag{1.12}$$

since $s(p|x)$ is of the form $s'(e_0|y)$, where $s'(q_0|y) = s(p^{\mathrm{lit}}|x)$.

Now, to prove (a), suppose first $p \sim_{E^a} q$ and $s(p^{\mathrm{lit}}|x) \in GT(E)$. Then e_0 does not occur in s. By (1.12), $s(p|x) \in GT(E^a)$, so $s(q^{\mathrm{lit}}|x) \in GT(E^a)$ by assumption, whence $s(q|x) \in GT(E)$ by (1.12).

In the other direction, suppose $p^{\mathrm{lit}} \sim_E q^{\mathrm{lit}}$ and $s(p|x) \in GT(E^a)$. This time s may contain e_0. But let s' be an e_0-free term such that $s = s'(e_0|y)$. Then $s(p|x) = s'(p, e_0|x, y) \in GT(E^a)$. So $s'(p, q_0|x, y) \in GT(E^a)$ by (1.10), and then $s'(p^{\mathrm{lit}}, q_0|x, y) \in GT(E)$ by (1.12). The assumption now yields $s'(q^{\mathrm{lit}}, q_0|x, y) \in GT(E)$, from which we get back to $s(q|x) \in GT(E^a)$.

This proves (a). The first part of (b) then follows (with $q = p^{\mathrm{lit}}$, since $p^{\mathrm{lit}} \sim_E p^{\mathrm{lit}}$); the second is clear from the above. (c) and (d) are easy consequences of (a). $\square$

By part (a) of this lemma, the extension to E^a preserves syntactic categories in the sense that for idiom-free terms p, q, $p \sim_E q$ implies $p \sim_{E^a} q$. By part (b), a grammatical term containing the idiom has the same syntactic category as its literal version, and the same surface form, just as we should expect. Parts (c) and (d) will be used in section 8.

To repeat, the atomic idiomatic extension E^a allows the idiom in exactly the same constructions as the literal version. So if $e_0 =$ *kick-the-bucket*, a sentence like *Mary feared that John had kicked the bucket* will presumably be ambiguous between an idiomatic and a literal reading, distinguished by the corresponding grammatical terms (recording their respective derivations).

What about *The bucket was kicked by John*? This depends on how the passive rule works in E. Assume as before that the literal version of *John kicked the bucket* is given by

$$\alpha(John, \alpha_0(kick, \gamma(the, bucket)))$$

and thus the idiomatic version by

$$\alpha(John, e_0).$$

If a passive rule β_p applies to *kick*, i.e., at the transitive verb, then *The bucket was kicked by John* would be given by something like

$$\alpha(\gamma(the, bucket), \alpha_0(\beta_p(kick), John)),$$

and there is no way to derive an idiomatic version of this, since the idiom *kick-the-bucket* will be an intransitive verb, to which passive does not apply. This treatment of *kick-the-bucket* as an atom is similar to the one in HPSG (cf. Sag and Wasow 1999, p. 269).

If, on the other hand, the passive rule of the grammar applies at or above the VP level, the literal version might be

$$\alpha_p(John, \alpha_0(kick, \gamma(the, bucket))).$$

Now an idiomatic version of *The bucket was kicked by John* is derivable in E^a, since the corresponding term

$$\alpha_p(John, e_0)$$

is *grammatical*. Whether it is *meaningful* or not is another matter. That depends on how the semantics of E is extended. We will returh to this issue at the end of section 8.

7 Paraphrase Semantics

We have to extend a given semantics μ for E. Actually, there are two ways to think about this; the first one is considered in this section. Assume then that e_0 has a *paraphrase* (for example, *die* for *kick the bucket*), by which I shall mean an expression $p_0 \in dom(\mu)$ such that $p_0 \sim_E q_0$ and $\mu(p_0) = m_0$ is the meaning we want to give to e_0.[7]

[7] Someone might object that *die* is not a paraphrase in this strict sense of *kick the bucket*, since, for example, *He was dying for five days* cannot be rendered using the idiom. But I am using the example only for illustration.

Since we have a paraphrase, each grammatical term $p = s(e_0|x)$ of E^a can be translated back into E; define

$$p^{\mathrm{tr}} = s(p_0|x).$$

By Lemma 6.1 (a) and the assumption that $p_0 \sim_E q_0$ it follows that $p^{\mathrm{tr}} \in GT(E)$, and also that $p_0 \sim_{E^a} e_0$.

Let

$$K^{\mathrm{tr}} = \{p \in GT(E^a) : p^{\mathrm{tr}} \in dom(\mu)\}.$$

That is, K^{tr} is the set of terms (possibly) containing the idiom whose paraphrases (replacing the idiom by its paraphrase) are meaningful. Now extend μ to K^{tr} in the obvious way:

$$\mu^{\mathrm{tr}}(p) = \mu(p^{\mathrm{tr}})$$

when $p \in K^{\mathrm{tr}}$; undefined otherwise. Clearly $\mu^{\mathrm{tr}}(e_0) = m_0$ and, for $p \in dom(\mu)$, $\mu^{\mathrm{tr}}(p) = \mu(p)$.

The proof of the next lemma is similar to the proof of Lemma 6.1 (a).

7.1 Lemma. *If $p, q \in GT(E^a)$ then*

(i) $p \sim_{E^a} q \iff p^{tr} \sim_E q^{tr}$

(ii) $p \sim_{\mu^{tr}} q \iff p^{tr} \sim_\mu q^{tr}$.

Now we can see exactly how μ^{tr} is related to μ.

7.2 Proposition. *Suppose that μ is compositional as in Definition 5.3, with meaning operations r_α corresponding to the function symbols $\alpha \in \Sigma$. Up to equivalence, μ^{tr} is the unique compositional extension of μ to K^{tr} such that $\mu^{tr}(e_0) = m_0$. Furthermore, μ^{tr} is the only semantics with these properties which has the* same *meaning operations as μ. In addition, if μ is (weakly) husserlian, so is μ^{tr}.*

Proof. Let ν be any semantics with the properties mentioned and let $p_i = s_i(e_0|x) \in K^{\mathrm{tr}}$, $i = 1, 2$. Then

$$\begin{aligned} p_1 \equiv_{\mu^{\mathrm{tr}}} p_2 \quad &\iff \quad s_1(p_0|x) \equiv_\mu s_2(p_0|x) \quad \text{(by definition of } \mu^{\mathrm{tr}}\text{)} \\ &\iff \quad s_1(p_0|x) \equiv_\nu s_2(p_0|x) \quad \text{(since } \nu \text{ extends } \mu\text{)} \\ &\iff \quad p_1 \equiv_\nu p_2. \end{aligned}$$

The last step follows using ν-compositionality twice (in the form given by Fact 5.4), since, by assumption, we have $e_0 \equiv_\nu p_0$. Hence, μ^{tr} and ν are equivalent in the sense of Definition 5.1.

Now, suppose $\alpha(p_1, \ldots, p_n) \in dom(\mu^{\mathrm{tr}})$. We claim that each p_i is in $dom(\mu^{\mathrm{tr}})$ and that

$$\mu^{\mathrm{tr}}(\alpha(p_1, \ldots, p_n)) = r_\alpha(\mu^{\mathrm{tr}}(p_1), \ldots, \mu^{\mathrm{tr}}(p_n)).$$

For, $\mu^{\mathrm{tr}}(\alpha(p_1, \ldots, p_n)) = \mu(\alpha(p_1, \ldots, p_n)^{\mathrm{tr}}) = \mu(\alpha(p_1^{\mathrm{tr}}, \ldots, p_n^{\mathrm{tr}})) = r_\alpha(\mu(p_1^{\mathrm{tr}}), \ldots, \mu(p_n^{\mathrm{tr}}))$. So $p_i^{\mathrm{tr}} \in dom(\mu)$, hence $p_i \in dom(\mu^{\mathrm{tr}})$, and the claim follows. Clearly, at most one semantics extending μ to K^{tr} and assigning m_0 to e_0 is determined in this way.

Finally, suppose $p \equiv_{\mu^{\mathrm{tr}}} q$. It follows that $p^{\mathrm{tr}} \equiv_\mu q^{\mathrm{tr}}$. If μ is (weakly) husserlian we have $p^{\mathrm{tr}} \sim_\mu q^{\mathrm{tr}}$ ($p^{\mathrm{tr}} \sim_E q^{\mathrm{tr}}$), and hence, by Lemma 7.1, $p \sim_{\mu^{\mathrm{tr}}} q$ ($p \sim_{E^a} q$). So μ^{tr} is (weakly) husserlian too. □

As Hodges (1998) stresses, mere compositionality is a rather weak requirement, and there are innumerable compositional extensions of μ to K^{tr} giving e_0 the meaning m_0. By the above proposition, they are all equivalent in the sense of having the same associated synonymy relation. But clearly, the paraphrase semantics μ^{tr} is the natural semantics here. Now recall that E and E^a have the same syntactic operations, and that compositionality means that there are meaning operations corresponding (homomorphically) to the syntactic ones. The proposition tells us that the natural semantics for E^a uses the *same* meaning operations as the given semantics for E.

Observe that μ^{tr} is not at all the *smallest* compositional idiomatic extension of μ assigning m_0 to e_0. Since e_0 is an atom, even $\mu' = \mu \cup \{\langle e_0, m_0\rangle\}$ preserves compositionality. But μ' is a trivial and uninteresting extension of μ, since it does not allow the idiom as a proper constituent in any meaningful term. It would be like adding the idiom *kick the bucket* but not allowing, say, *John kicked the bucket* to be read idiomatically. Instead we chose K^{tr} as the new domain. The reasonableness of this choice is discussed further in the following sections.

8 New Idiomatic Meanings

The strategy from the preceding section does not work if there is no expression of the same syntactic category with exactly the desired idiomatic meaning, a situation which might be quite common. There is still a clear intuition, I think, that one should be able to easily modify the existing compositional machinery to accommodate the idiom. This intuition can be cashed out in the following way. Assume that E is categorial (Definition 5.5). We now slightly strengthen the assumption of compositionality.

8.1 Definition.

- A *type system* for E has the form $T = \langle r_\alpha, D_c \rangle_{\alpha \in \Sigma, c \in Cat_E}$. $D_{[p]_E}$ is the *type* of p, and we assume that types are pairwise disjoint, and that for $\alpha \in \Sigma$, if $\alpha(p_1, \ldots, p_n) \in GT(E)$, r_α is a *total* function

$$r_\alpha : D_{[p_1]_E} \times \cdots \times D_{[p_n]_E} \longrightarrow D_{[\alpha(p_1,\ldots,p_n)]_E}$$

(E is categorial, so this is independent of the choice of $p_1, \ldots, p_n$).

- T is *μ-compositional* if $dom(\mu)$ is closed under subterms and, whenever $p, \alpha(p_1, \ldots, p_n) \in dom(\mu)$,

  (i) $\mu(p) \in D_{[p]_E}$

  (ii) $\mu(\alpha(p_1, \ldots, p_n)) = r_\alpha(\mu(p_1), \ldots, \mu(p_n))$.[8]

Actually, this just adds a familar property to compositionality:

8.2 Proposition. *Let μ be a semantics for a categorial E. The following are equivalent:*

(a) *μ is weakly husserlian and compositional.*

(b) *There is a μ-compositional type system for E.*

Proof. To verify that if T is a μ-compositional type system for E then μ is weakly husserlian, assume $p \equiv_\mu q$. Then $\mu(p) = \mu(q)$. Since $\mu(p) \in D_{[p]_E}$ and $\mu(q) \in D_{[q]_E}$, it follows that $D_{[p]_E} = D_{[q]_E}$, so $p \sim_E q$.

In the other direction, let μ be weakly husserlian and compositional, with functions r_α corresponding to the $\alpha \in \Sigma$. We may assume that $dom(r_\alpha)$ consists of just those $(m_1, \ldots, m_n)$ such that there are $p_1, \ldots, p_n$ with $m_i = \mu(p_i)$ and $\alpha(p_1, \ldots, p_n) \in dom(\mu)$. Define, for each $p \in GT(E)$,

$$D_{[p]_E} = \{\mu(q) : q \in [p]_E \cap dom(\mu)\}.$$

It follows that if $c_1 \neq c_2$ then $D_{c_1} \cap D_{c_2} = \emptyset$. For, suppose $c_i = [p_i]_E$ and $m \in D_{c_1} \cap D_{c_2}$. Then there are q_1, q_2 such that $m = \mu(q_1) = \mu(q_2)$ and $p_i \sim_E q_i$, $i = 1, 2$. So $q_1 \equiv_\mu q_2$, hence $q_1 \sim_E q_2$ since μ is weakly husserlian. Thus, $p_1 \sim_E p_2$, i.e., $c_1 = c_2$.

It remains to extend each r_α to a total function r'_α. But this trivial: just choose an arbitrary object of the right type when the argument

[8] In condition (i), it suffices to assume $\mu(a) \in D_{[a]_E}$ for atomic $a \in dom(\mu)$.

is not in the domain of r_α. It is then straightforward to check that $\langle r'_\alpha, D_c\rangle_{\alpha\in\Sigma, c\in Cat_E}$ is a μ-compositional type system for E. $\square$

Assume, then, that T is a μ-compositional type system for E. By Lemma 6.1 (c) and (d) it follows that $T = \langle r_\alpha, D_{c'}\rangle_{\alpha\in\Sigma, c\in Cat_E}$, where $D_{c'} = D_c$, is also a type system for E^a. Now we can use T to extend μ inductively to a semantics μ^a such that $\mu^a(e_0) = m_0$, even if the desired meaning m_0 for our idiom is not in the range of μ; it suffices that it is in $D_{[q_0]_E}$. For the meaning of e_0 is given by an atomic stipulation, and the extension to complex terms is taken care of by the (total) meaning operations of T.

What should the domain of μ^a be? We have no paraphrase this time — all we have is the literal reading q_0. Therefore, an obvious choice is to let $dom(\mu^a)$ be

$$K^{\text{lit}} = \{p \in GT(E^a) : p^{\text{lit}} \in dom(\mu)\},$$

the set of terms (possibly) containing the idiom whose literal versions (defined earlier, just before Lemma 6.1) are meaningful.

Thus we define μ^a inductively as follows, making sure that at each step $p \in dom(\mu^a) \Leftrightarrow p^{\text{lit}} \in dom(\mu)$:

- $\mu^a(e_0) = m_0$, and $\mu^a(a) = \mu(a)$ for $a \in A$ (whenever defined).
- Let $\alpha(p_1, \ldots, p_n) \in GT(E^a)$. If $\alpha(p_1, \ldots, p_n)^{\text{lit}} = \alpha(p_1^{\text{lit}}, \ldots, p_n^{\text{lit}})$ is in $dom(\mu)$ then so is each p_i^{lit}, so $\mu^a(p_i)$ is defined, by induction hypothesis, and we let

  $\mu^a(\alpha(p_1, \ldots, p_n)) = r_\alpha(\mu^a(p_1), \ldots, \mu^a(p_n))$

  (undefined otherwise). Note that this works since the r_α are total.

The following theorem summarizes most of our findings so far.

8.3 Theorem. *Suppose that E is categorial and that E^a is obtained by adding $e_0 \in E - A$ as a new atom, where $e_0 = val(q_0)$. Suppose further that μ is a semantics for E in which q_0 is meaningful, and that $T = \langle r_\alpha, D_c\rangle_{\alpha\in\Sigma, c\in Cat_E}$ is a μ-compositional type system for E. Finally, suppose that $m_0 \in D_{[q_0]_E}$. Then the following holds:*

(a) *There is a unique extension μ^a of μ to K^{lit} such that T is μ^a-compositional and $\mu^a(e_0) = m_0$.*

(b) *Suppose $m_0 = \mu(p_0)$ for some paraphrase $p_0 \in dom(\mu)$ such that $p_0 \sim_E q_0$. Then for all $p \in K^{tr} \cap K^{lit}$, $\mu^{tr}(p) = \mu^a(p)$.*

Proof. As for (a), it is straightforward to verify that μ^a has the required properties, and since this semantics is determined by the meaning operations of T, it is unique. For (b), suppose $p \in K^{\text{tr}} \cap K^{\text{lit}}$. Then

$$\begin{array}{rcll} \mu^{\text{tr}}(p) & = & \mu(p^{\text{tr}}) & (\text{since } p \in \mathit{dom}(\mu^{\text{tr}})) \\ & = & \mu^a(p^{\text{tr}}) & (\text{since } \mu^a \text{ extends } \mu) \\ & = & \mu^a(p) & \end{array}$$

The last step follows by μ^a-compositionality (Fact 5.4), since $e_0 \equiv_{\mu^a} p_0$ and $p, p^{\text{tr}} \in \mathit{dom}(\mu^a)$. $\square$

When we have a paraphrase p_0 of the idiom, there are two ways of extending the semantics to E^a; by (b) above we can rest assured that they never give conflicting results.

Note that if $p_0 \sim_\mu q_0$, then $K^{\text{tr}} = K^{\text{lit}}$. However, in general there is no reason to suppose this to be the case. As we indicated at the end of section 6, one way of treating the passive operation results in an idiomatic version of *The bucket was kicked by John*, recorded by

$$\alpha_p(\mathit{John}, e_0),$$

where $e_0 = \mathit{kick\text{-}the\text{-}bucket}$. (For the other, perhaps more plausible, way of treating the passive there was no idiomatic version of this sentence.) Thus, $\alpha_p(\mathit{John}, e_0) \in K^{\text{lit}}$, since the literal version of *The bucket was kicked by John* is μ-meaningful. By assumption, $\mathit{die} \sim_{E^a} e_0$, and therefore $\alpha_p(\mathit{John}, \mathit{die})$ is grammatical.[9] But nothing forces us to regard the latter term as meaningful; indeed one might suppose that any respectable μ rules it out as meaningless. If so, $\alpha_p(\mathit{John}, e_0) \notin K^{\text{tr}}$, so $K^{\text{lit}} \not\subseteq K^{\text{tr}}$.

For an example showing that the inclusion in the other direction can fail as well, we might consider, say, *John died of cancer*; if a literal reading of *John kicked the bucket of cancer* is meaningless it would follow that $K^{\text{tr}} \not\subseteq K^{\text{lit}}$.[10]

9 Idioms with Structure 1

Now suppose instead we want to consider (some or all) idioms as having syntactic structure. It seems reasonable that the idiomatic reading has

[9] As to the (surface) value of $\alpha_p(\mathit{John}, \mathit{die})$, such a grammar might stipulate, for example, that it is the same as that of $\alpha(\mathit{John}, \mathit{die})$, i.e., *John died.*

[10] Assuming, perhaps dubiously, that the idiomatic reading of that sentence is meaningful.

the *same* structure as the literal $q_0 = \alpha_0(q_{01}, \dots q_{0k})$. But we still need to distinguish them to avoid ambiguity, while respecting the intuition that at surface level the two versions coincide. Here is the perhaps simplest way to achieve this: we keep using the old operation $\underline{\alpha_0}$, but give it a new name. This means that our given grammar E is extended in the following way.

9.1 Definition.

$$E^i = (E, A, \underline{\alpha})_{\alpha \in \Sigma^i},$$

where $\Sigma^i = \Sigma \cup \{\alpha_0^i\}$, and α_0^i is a new k-ary function symbol such that $\underline{\alpha_0^i} = \underline{\alpha_0}$. E^i is called a *duplicated rule extension* of E. Let $q_0^i = \alpha_0^i(q_{01}, \dots, q_{0k})$.

So now we have a new function symbol, hence new grammatical terms, but the same expressions and the same operations on expressions as before. For example, the sentence *John kicked the bucket* could have a derivation recorded by the term

$$\alpha(\mathit{John}, \alpha_0^i(\mathit{kick}, \gamma(\mathit{the}, \mathit{bucket})))$$

in addition to the ordinary one. The syntactic structure of the idiom, and the way it combines with larger structures, is exactly the same as for the literal version. In principle, all we have done is to place a 'marker' on the (derivation of the) idiom.

The relations between E and E^i are easily described. To obtain the literal version of a term this time we only need to replace 'α_0^i' by 'α_0'. Thus, for $s \in T(E^i)$, define

$$s^- = \text{ the result of deleting all superscripts } {}^i \text{ in } s.$$

Note that $s^{--} = s^-$ and $s(p|x)^- = s^-(p^-|x)$. Also define, for $X \subseteq T(E)$,

$$X^+ = \{s \in T(E^i) : s^- \in X\}.$$

That is, X^+ consists of all those terms that are obtained from terms in X by adding the superscript ^{i} to zero or more occurrences of 'α_0'.

9.2 Lemma.

(a) *$T(E^i) = T(E)^+$, $GT(E^i) = GT(E)^+$, and $val^i(p) = val(p^-)$ for $p \in GT(E^i)$.*

(b) *For* $p, q \in GT(E^i)$, $p \sim_{E^i} q \Leftrightarrow p^- \sim_E q^-$.

(c) *If* $p \in GT(E^i)$ *then* $[p]_{E^i} = [p^-]_{E^i}$, *and if* $p \in GT(E)$ *then* $[p]_{E^i} = [p]^+_E$.

(d) *If* E *is categorial, so is* E^i, *and the map* $'$ *from* Cat_E *to* Cat_{E^i} *defined by* $([p]_E)' = [p]_{E^i}$ *is a bijection.*

Proof. (a) is straightforward from the definition of E^i, since the terms $\alpha_0(p_1, \ldots, p_k)$ and $\alpha_0^i(p_1, \ldots, p_k)$ are grammatical under exactly the same circumstances. To prove (b), suppose first $p \sim_{E^i} q$ and $s(p^-|x) \in GT(E)$, where $s \in T(E)$. Then $s^- = s$, so $s(p^-|x) = s(p|x)^-$, and hence $s(p|x) \in GT(E^i)$ (by (a)). Therefore $s(q|x) \in GT(E^i)$ by assumption, and it follows that $s(q^-|x) \in GT(E)$. In the other direction, suppose $p^- \sim_E q^-$ and $s(p|x) \in GT(E^i)$, where $s \in T(E^i)$. Then $s(p|x)^- = s^-(p^-|x) \in GT(E)$, so $s^-(q^-|x) \in GT(E)$, from which we deduce $s(q|x) \in GT(E^i)$. This shows (b). (c) and (d) follow from (b). □

As to semantics, the reason we can still obtain compositionality is that two distinct semantic operations can correspond to α and α_0^i. Again, there are two ways to extend the semantics μ. If there is a paraphrase, i.e., a term p_0 such that $p_0 \sim_E q_0$ and $\mu(p_0) = m_0$, then we can define the new semantics directly by means of translation as before: For $p \in GT(E^i)$, let $p^\dagger$ be the result of replacing each occurrence of q_0^i by p_0, and let $\mu^\dagger(p) = \mu(p^\dagger)$ whenever the latter is defined. Thus the domain of $\mu^\dagger$ is

$$K^\dagger = \{s(q_0^i|x) : s(p_0|x) \in dom(\mu)\}.$$

We can now prove results corresponding to those in section 7 for this semantics.

Since the idiom has structure, we can insert other phrases in that structure than the original ones. For example, we can derive (*John lifted the bucket*)^i, by which I here mean the surface string — identical to its non-idiomatic correspondent — which is the value of the term

$$\alpha(John, \alpha_0^i(lift, \gamma(the, bucket))).$$

Likewise, we get (*John kicked the pail*)^i, corresponding to

$$\alpha(John, \alpha_0^i(kick, \gamma(the, pail))).$$

By the definition of E^i, both of these are grammatical (since the results of deleting ^{i} are grammatical).

However, in the paraphrase semantics, none of these terms is meaningful, since none of them has the idiom $q_0^i = \alpha_0^i(kick, \gamma(the, bucket))$ as a constituent. So the paraphrase semantics avoids this kind of overgeneration.

The paraphrase semantics also avoids idiomatic versions of *The bucket was kicked by John.* Two ways of treating the passive were mentioned at the end of section 6. In the present case we obtain

$$\alpha(\gamma(the, bucket), \alpha_0^i(\beta_p(kick), John)) \tag{1.13}$$

and

$$\alpha_p(John, \alpha_0^i(kick, \gamma(the, bucket))), \tag{1.14}$$

respectively. Both of these are grammatical in E^i, but (1.13) has no translation into E, and we saw at the end of the previous section that the translation of (1.14), $\alpha_p(John, die)$, can safely be assumed to be μ-meaningless.

On the other hand, one might argue that it is more natural to define the semantics for E^i by the usual induction over syntax instead. And one might also claim there is no great harm in allowing 'idiomatic readings' of $(John\ lifted\ the\ bucket)^i$ and $(John\ kicked\ the\ pail)^i$, as long as the former is the same as the literal reading of that sentence, and the latter means that John died.[11] Or, that those readings may be ruled out by some different mechanism. The second way of extending the semantics allows for this.

Suppose, then, that E is categorial, and that $T = \langle r_\alpha, D_c \rangle_{\alpha \in \Sigma, c \in Cat_E}$ is a μ-compositional type system for E. Assume $m_0 \in D_{[q_0]_E}$. Letting $D_{c'} = D_c$ as before, define a new type system

$$T^i = \langle r_\alpha, D_d \rangle_{\alpha \in \Sigma^i, d \in Cat_{E^i}},$$

where $r_{\alpha_0^i}$ is a function with the same domain as r_{α_0}, defined as follows:

$$r_{\alpha_0^i}(m_1, \ldots, m_k) = \begin{cases} m_0 & \text{if } m_j = \mu(q_{0j}), 1 \leq j \leq k \\ r_{\alpha_0}(m_1, \ldots, m_k) & \text{otherwise.} \end{cases}$$

If $\alpha_0^i(p_1, \ldots, p_k) \in GT(E^i)$, $(\alpha_0^i(p_1, \ldots, p_k))^- = \alpha_0(p_1^-, \ldots, p_k^-)) \in GT(E)$. Also, $D_{[p_j^-]_E} = D_{[p_j]_{E^i}}$ and $D_{[\alpha_0(p_1^-, \ldots, p_k^-)]_E} = D_{[\alpha_0^i(p_1, \ldots, p_k)]_{E^i}} = D_{[q_0]_{E^i}}$ (since E^i is categorial). Hence,

$$r_{\alpha_0^i} : D_{[p_1]_{E^i}} \times \cdots \times D_{[p_k]_{E^i}} \longrightarrow D_{[\alpha_0^i(p_1, \ldots, p_k)]_{E^i}}.$$

[11] It *has to* mean this if compositionality is not to be violated.

It follows that T^i is a type system for E^i; essentially it is T plus the new semantic operation corresponding to α_0^i. So we can use T:s compositional machinery to extend μ inductively to a semantics μ^i for E^i such that $\mu^i(q_0^i) = m_0$, making sure at each step that $p \in dom(\mu^i) \Leftrightarrow p^- \in dom(\mu)$:

- $\mu^i(a) = \mu(a)$ for $a \in A$ (whenever defined).
- Let $p = \alpha(p_1, \ldots, p_n)$ be a complex term in $GT(E^i)$. p^- is of the form $\beta(p_1^-, \ldots, p_n^-)$, where β is α if $\alpha \in \Sigma$, and β is α_0 if $\alpha = \alpha_0^i$. If $p^-$is in $dom(\mu)$ then so is each p_j^-, so $\mu^i(p_j)$ is defined, by induction hypothesis, and we let

  $\mu^i(p) = r_\alpha(\mu^i(p_1), \ldots, \mu^i(p_n))$ (undefined otherwise).

We get $dom(\mu^i) = K^i = dom(\mu)^+$. The proof of the next theorem is similar to the proof of Theorem 8.3.

9.3 Theorem. *Suppose that E is categorial and that E^i is obtained as above by duplicating the last syntactic rule used in forming the term q_0. Suppose further that $T = \langle r_\alpha, D_c \rangle_{\alpha \in \Sigma, c \in Cat_E}$ is a μ-compositional type system for E, and that we wish to give q_0 the idiomatic meaning $m_0 \in D_{[q_0]_E}$. Then the following holds:*

(a) *There is a unique extension μ^i of μ to K^i such that T^i is μ^i-compositional (and hence $\mu^i(q_0^i) = m_0$).*

(b) *Suppose $m_0 = \mu(p_0)$ for some paraphrase $p_0 \in dom(\mu)$ such that $p_0 \sim_E q_0$, and suppose μ is extended to $\mu^\dagger$ as described above. Then for all $p \in K^\dagger \cap K^i$, $\mu^\dagger(p) = \mu^i(p)$.*

If $p_0 \sim_\mu q_0$ then $K^\dagger \subseteq K^i$, but as we saw, in general this cannot be assumed.

Now we get the announced result for the meanings of (*John lifted the bucket*)^i and (*John kicked the pail*)^i. A little more abstractly, recalling that $q_0^i = \alpha_0^i(q_{01}, \ldots, q_{0k})$, let q be a grammatical term of E of the same syntactic category as q_{01}. It follows that if q is *not* synonymous with q_{01} (as *lift* is not synonymous with *kick*), then

$$\mu^i(\alpha_0^i(q, q_{02}, \ldots, q_{0k})) \;=\; \mu(\alpha_0(q, q_{02}, \ldots, q_{0k})).$$

But if $q \equiv_\mu q_{01}$ (as perhaps *bucket* $\equiv_\mu$ *pail*), then

$$\mu^i(\alpha_0^i(q, q_{02}, \ldots, q_{0k})) \;=\; m_0.$$

As to (*The bucket was kicked by John*)^i, the μ^i-meaning of (1.13) is the literal one, and the μ^i-meaning of (1.14) is $r_{\alpha_p}(\mu(John), m_0)$. This is defined because r_{α_p} is assumed to be total, but the value might be arbitrarily stipulated.[12]

In view of these observations, in particular the last one (which could however also be taken as an argument against that analysis of the passive), one might argue that the price paid here for simplicity is too high, and that a proper idiomatic extension should not allow such terms as meaningful. A particular theory might have the means to *filter out* some terms of K^i, thus settling for a semantics μ' between μ and μ^i. Of course, when there is a paraphrase, $\mu^\dagger$ is such a semantics. We end this section by noting that *any* such semantics is in a sense already covered by the above result:

9.4 Corollary. *Let μ' be any semantics such that (i) $\mu \subseteq \mu' \subseteq \mu^i$, (ii) $K' = dom(\mu')$ is closed under subterms, and (iii) $q_0^i \in K'$. Then (a) and (b) of Theorem 9.3 hold with μ' and K' in place of μ^i and K^i.*

Proof. We only check compositionality: if $\alpha(p_1, \ldots, p_n)$ is μ'-meaningful, then $\mu'(\alpha(p_1, \ldots, p_n)) = \mu^i(\alpha(p_1, \ldots, p_n)) = r_\alpha(\mu^i(p_1), \ldots, \mu^i(p_n)) = r_\alpha(\mu'(p_1), \ldots, \mu'(p_n))$, since K' is closed under subterms. □

10 Idioms with Structure 2

The extensions of (E, μ) to (E^a, μ^a) and to (E^i, μ^i) are strikingly simple, but there is one intuition that neither account captures. This is the idea, explicit in Nunberg et al. (1994), that an idiom like *pull strings* not only has the same syntactic structure as the literal version, but that its meaning is obtained by the *same meaning operation* as in the literal case, only that this operation is now applied to new idiomatic meanings of *pull* and *strings*.

Indeed, thinking for the moment of *pull strings* as $\alpha_0(pull, strings)$, merely introducing a new meaning operation $r_{\alpha_0^i}$ as in the previous section seems less suitable here, since this idiom occurs in much more varied constructions than *kick the bucket*. We have idiomatic readings of, for example, *Strings were pulled to get John his position*, and *Mary tried to pull a lot of strings*. Thus, rather than defining several new

[12] One could argue that for this treatment of the passive it is natural to have r_α and r_{α_p} give the same value when both are defined, and so $r_\alpha = r_{\alpha_p}$ in the total case. But then we get the idiomatic reading *John died* of (*The bucket was kicked by John*)^i, which also is not really what we want.

meaning operations, it seems more adequate to introduce new atoms $pull^i$ and $strings^i$ instead and use the old syntactic and semantic operations, if possible.

The notion of an atomic extension from section 6 is not directly applicable here. This is because, in the first place, there are no available expressions in E and, in the second place, even if we could add such atoms, there would be no difference at the surface level between *pull* and $pull^i$, whereas we need to distinguish between them as grammatical terms, to avoid ambiguity.[13]

But it is easy to adapt the framework to handle these problems, assuming, as above, that the 'idiom chunks' to be added are all atoms.

The change we need is minimal: let a *grammar* have the form

$$((E, A, \underline{\alpha})_{\alpha\in\Sigma}, v),$$

where E and Σ are as before, but we no longer assume that $A \subseteq E$; instead we have a function v from A to E.[14] This idea is in fact not at all *ad hoc*: to account for *lexical ambiguity* (or homonymity) we need in any case to be able to distinguish between an atom as a grammatical term and its surface form. Thus, for example, we could have distinct atoms $bank_1$ and $bank_2$ in A, but $v(bank_1) = v(bank_2) \in E$.

Very few changes to Definition 5.1 are needed:

- In the definition of $T(E)$, let $A \cup X \subseteq T(E)$ for the base case; the case of complex terms is as before.
- In the definition of $GT(E)$ and val, the base case is
  - $a \in A$ is an *atomic* grammatical term, and $val(a) = v(a)$;

  the case of complex terms is again as before.

All the other notions (and results) from section 5 are exactly as before.

Now, assume that μ is a compositional semantics for E, with meaning operations r_α corresponding to α, for $\alpha \in \Sigma$. The presupposition in the present case is that there are new atoms $a^i_1, \ldots, a^i_k$ such that

$$v(a^i_j) = v(a_j),\ 1 \leq j \leq k,$$

and further that there are meanings $m^i_1, \ldots, m^i_k$ such that m^i_j is the meaning of a^i_j, $1 \leq j \leq k$, and, roughly, the meaning operations of μ

[13] This problem did not arise before, precisely because e_0 was the value of a *complex* term.

[14] So our previous notion of grammar is the special case when $A \subseteq E$ and v is the identity function.

yield idiomatic readings of certain expressions when applied to the m^i_j (or to arguments obtained by applying other meaning operations to the m^i_j). Note that in this case it is less natural to single out as before a particular term as *the* idiom (say, $q^i_0 = \alpha_0(a^i_1, \dots, a^i_k)$), since several forms of expression can have idiomatic readings.

So the idiom extension problem is how to extend E and μ in a natural way to cover this.

In fact, one way to proceed should by now be fairly obvious. First, we extend E to

$$E^* = ((E, A \cup \{a^i_1, \dots, a^i_k\}, \underline{\alpha})_{\alpha\in\Sigma}, v^*),$$

where $v*$ extends v and $v^*(a^i_j) = v(a_j)$, $1 \le j \le k$. Note that there is no need to redefine the syntactic rules, since the surface forms have not changed. So for example, if *Mary tried to pull several strings* has a derivation in E reflected by the term

$$\alpha(\mathit{Mary}, \alpha_0(\delta(\mathit{try\text{-}to}, \mathit{pull}), \gamma(\mathit{several}, \mathit{strings}))), \tag{1.15}$$

then the idiomatic version

$$\alpha(\mathit{Mary}, \alpha_0(\delta(\mathit{try\text{-}to}, \mathit{pull}^i), \gamma(\mathit{several}, \mathit{strings}^i))), \tag{1.16}$$

is grammatical in E^* provided $v^*(\mathit{pull}^i) = v(\mathit{pull})$, and $v^*(\mathit{strings}^i) = v(\mathit{strings})$.

As in section 9, we let s^- be the result of deleting all superscripts ^{i} in s (though now we delete them from the new atoms, not from a new function symbol). Again, we have $s^{--} = s^-$ and $s(p|x)^- = s^-(p^-|x)$, which means that the following lemma is proved almost exactly as Lemma 9.2:

10.1 Lemma.

(a) *$T(E^*) = T(E)^+$, $GT(E^*) = GT(E)^+$, and $val^*(p) = val(p^-)$ for $p \in GT(E^*)$.*

(b) *For $p, q \in GT(E^*)$, $p \sim_{E^*} q \Leftrightarrow p^- \sim_E q^-$.*

(c) *If $p \in GT(E^*)$ then $[p]_{E^*} = [p^-]_{E^*}$, and if $p \in GT(E)$ then $[p]_{E^*} = [p]^+_E$.*

(d) *If E is categorial, so is E^*, and the map $'$ from Cat_E to Cat_{E^*} defined by $([p]_E)' = [p]_{E^*}$ is a bijection.*

Consider first the paraphrase semantics. In this case it presupposes that there are terms $p_j \in dom(\mu)$ such that $a_j \sim_E p_j$ and $\mu(p_j) = m^i_j$, $1 \leq j \leq k$. Again we can translate back to E in the obvious way: Each $p \in GT(E^*)$ equals $s(a^i_1, \ldots, a^i_k | x_1, \ldots, x_k)$ for a unique $s \in T(E)$; let

$$p^{\mathrm{tr}^*} = s(p_1, \ldots, p_k | x_1, \ldots, x_k).$$

Thus, $(a^i_j)^{\mathrm{tr}^*} = p_j$ and, if $p \in GT(E)$, $p^{\mathrm{tr}^*} = p$. As in section 7 we get a translation semantics by putting

$$\mu^{\mathrm{tr}^*}(p) = \mu(p^{\mathrm{tr}^*})$$

whenever this is defined. So $\mu^{\mathrm{tr}^*}(a^i_j) = m^i_j$, and if $p \in dom(\mu)$, then $\mu^{\mathrm{tr}^*}(p) = \mu(p)$.

Now, just as in the proof of Proposition 7.2, we can show that if $\alpha(q_1, \ldots, q_n) \in dom(\mu^{\mathrm{tr}^*})$ then each q_j is in $dom(\mu^{\mathrm{tr}^*})$ and

$$\mu^{\mathrm{tr}^*}(\alpha(q_1, \ldots, q_n)) = r_\alpha(\mu^{\mathrm{tr}^*}(q_1), \ldots, \mu^{\mathrm{tr}^*}(q_n)).$$

So μ^{tr^*} is a compositional extension of μ with the same meaning operations. (The rest of Proposition 7.2 also carries over.)

Thus, assuming just for the sake of the argument here that *pull*^{i} paraphrases as *utilize*, and *strings*^{i} as *connections*, we get the desired reading of (1.16). But what about

$$\alpha(\mathit{Mary}, \alpha_0(\delta(\mathit{try\text{-}to}, \mathit{pull}), \gamma(\mathit{several}, \mathit{strings}^i))) \tag{1.17}$$

(which has the same surface value), or

$$\alpha(\mathit{Mary}, \alpha_0(\delta(\mathit{try\text{-}to}, \mathit{tie}), \gamma(\mathit{several}, \mathit{strings}^i))) \tag{1.18}$$

(with the value *Mary tried to tie several strings*)? Both are grammatical in E^* (provided, in the latter case, $\mathit{tie} \sim_E \mathit{pull}$). But presumably none is μ^{tr^*}-meaningful. This is because there is every reason to suppose that *strings* and *connections* are not of the same semantic category, and indeed that *Mary tried to pull several connections* and *Mary tried to tie several connections* are not meaningful.

However, in general we cannot assume that there are paraphrases (indeed Nunberg et al. (1994) appear to claim that there almost never are). So suppose that E is categorial and that $T = \langle r_\alpha, D_c \rangle_{\alpha \in \Sigma, c \in Cat_E}$ is a μ-compositional type system for E such that $m^i_j \in D_{[a_j]_E}$, $1 \leq j \leq k$. By the previous lemma, with $D_{c'} = D_c$, $T = \langle r_\alpha, D_{c'} \rangle_{\alpha \in \Sigma, c \in Cat_E}$ is also a type system for E^*. Now it is obvious how to extend μ to a semantics μ^* with domain $K^* = \{p : p^- \in dom(\mu)\}$: add $\mu^*(a^i_j) = m^i_j$ to the atomic clauses, and define

$$\mu^*(\alpha(q_1, \dots, q_n)) = r_\alpha(\mu^*(q_1), \dots, \mu^*(q_n))$$

when $\alpha(q_1, \dots, q_n)^- \in dom(\mu)$. The analogue of Theorem 9.3 holds.

But this semantics risks some serious overgeneration. For example, it will find (1.18) meaningful (because *Mary tried to tie several strings* is μ-meaningful), and, if m^i_1 is the meaning of *strings*^{i}, its meaning will be

$$m \;=\; r_\alpha(\mu(Mary), r_{\alpha_0}(r_\delta(\mu(try\text{-}to), \mu(tie)), r_\gamma(\mu(several), m^i_1))).$$

The meaning operations are assumed to be total, so *some* meaning m of (1.18) is obtained, but we have no 'control' over what m is. And just stipulating does not seem to help; there simply is no sensible candidate.

The problem here is worse than for the semantics μ^i in section 9.[15] There we had a new meaning operation $r_{\alpha^i_0}$ which coincided with r_{α_0} except in one case, namely, when the arguments were the ones given by the 'parts' of the idiom. Therefore (*John lifted the bucket*)^i was meaningful, but it just meant that John lifted the bucket. Likewise, (*John kicked the pail*)^i would mean that John died, if *pail* is synonymous with *bucket.* These are overgenerations, to be sure, but relatively harmless compared to *Mary tried to tie several strings*^{i}.

So one needs to somehow filter out unwanted terms like (1.18) as meaningless. Particular semantic theories might have various ways of doing this — cf. also the discussion in the next section. For example, one could stipulate a constraint of roughly the following kind (using the notation of the previous examples):

(1.19) A term in $GT(E^*)$ is meaningful iff it belongs to K^* and for each subterm of the form $\alpha(p, \alpha_0(q_1, q_2))$ it holds that *pull*^{i} occurs in q_1 iff *strings*^{i} occurs in either q_2 or p.

This would rule out (1.17) and (1.18), but allow (1.16) as well as an idiomatic reading of *Several strings were pulled by Mary.*

A different approach could be used if, although there are no paraphrases of the a^i_j, there exist μ-meaningful terms p'_j such that $a_j \sim_E p'_j$, which provide the semantic categories of $a^i_1, \dots, a^i_k$ in the extended semantics. A term in K^* would then be stipulated to be meaningful only if the result of replacing all the a^i_j occurring in it by p'_j is μ-meaningful.

[15] And, of course, for the semantics μ^a of section 8, where the problem could not even arise, since the idiom had no structure.

This would rule out unwanted terms in a way similar to the paraphrase semantics, even in the absence of paraphrases.

From the abstract algebraic point of view taken in this paper, it seems sufficient to point out that *however* such constraints on the semantics are carried out, compositionality is not in danger. For, in analogy with Corollary 9.4, we get

10.2 Corollary. *Let μ' be any semantics such that (i) $\mu \subseteq \mu' \subseteq \mu^*$, (ii) $dom(\mu')$ is closed under subterms, and (iii) $\mu'(a^i_j) = m^i_j$, $1 \leq j \leq k$. Then μ' is compositional, with the given r_α as meaning operations.*

11 Discussion

We have looked at some ways of extending both syntax and semantics in order to incorporate a new idiom. To see that these extensions are at least reasonable, one may observe that criteria like the following have been satisfied:

- *Surface identity.* On the surface, the idiom looks the same as the non-idiomatic expression.
- *Preservation.* The extensions preserve various desirable properties, in particular, compositionality. We saw that other properties (such as the husserl property) were preserved too, and one may also verify (where E' and μ' stand for the extensions considered here) that for $p, q \in GT(E)$,
  - $p \sim_E q$ implies $p \sim_{E'} q$ (preservation of syntactic categories),
  - $p \sim_\mu q$ implies $p \sim_{\mu'} q$ (preservation of semantic categories).
- *Uniqueness.* The extension is (in some suitable sense) uniquely determined (given that it should have certain properties).
- *Conservativity.* If $p \in GT(E) - dom(\mu)$ then $p \in GT(E') - dom(\mu')$. I have not discussed this property, but it does hold, and clearly it is a reasonable one.
- *Distribution.* The extended semantics should have a reasonable and non-trivial domain, allowing the idiom in all contexts one expects to be able to use but, ideally, not in others.

In connection with the issue of uniqueness one may note that in all cases considered, the extended semantics used (essentially) the same operations of meaning composition as the original semantics. This is intuitively satisfactory, and emphasizes the fact that we are not merely interested in the existence of a compositional extension, but that there is a unique *natural* extension, given that we have chosen to analyze the idiom in one of the ways proposed here.

The question of an adequate distribution highlights the issue of *overgeneration*. Examples of two kinds have been considered throughout this paper. The passive construction represents the first kind: some idioms 'passivize', others don't. An atomic analysis of idioms like *kick the bucket* may prevent passivization. With a different analysis of the passive, however, it doesn't. But even so, and even with a non-atomic analysis, we saw that the paraphrase semantics allows transfer of semantic restrictions (such as the constraint that passivization only make sense for 2-place predicates) to the idiomatic extension, thus blocking overgeneration. On the other hand, when there is no paraphrase, the most straightforward idiomatic extensions did allow an idiomatic reading of *The bucket was kicked by John* as *John died*, a reading which would then have to be ruled out in some other way.

The second kind of example concerns cases when 'ordinary' expressions were inserted in idiomatic constructions, or, to put it differently, when some parts of the idiom were missing. This can only happen with a non-atomic analysis of the idiom. There were some not so serious cases, as when *John lifted the bucket* had an idiomatic syntax but the usual meaning, or, more seriously, when *John kicked the pail* meant that John died (assuming that *pail* and *bucket* are synonymous), or, even more seriously, when *Mary tried to pull several strings* was formed with an idiomatic meaning of *strings*, which resulted in the sentence making no sense at all.

We saw that when paraphrases are available, these putative examples could usually be ruled out. But in general one cannot assume that idioms can be paraphrased (in the strict sense used here). Then some other mechanism will be needed to filter out the unwanted readings. Two indications were given at the end of section 10 of how this might be done. But one should not really expect our abstract algebraic framework to be able to describe such mechanisms in detail. What we could prove, however, was that however they are enforced, the idiomatic semantic extension will still be compositional, and moreover use the same meaning operations as the given semantics (Corollary 10.2).

One feature of the approach taken here is the distinction between syntactic wellformedness and meaningfulness. This is partly for practical reasons. I wanted the various idiomatic extensions of the syntax to be describable within the abstract framework, and have a precise and transparent relation to the old syntax. They provide, so to speak, a maximal range of terms, which can then be restricted in various ways to a set of meaningful terms. Some such ways can be characterized at the present level of generality, others not. Since most of the results here about the compositionality of idiomatic semantic extensions merely assume that the semantic domain is a (reasonable) subset of the set of grammatical terms, they will apply to more detailed ways of specifying that subset as well. Also, I could in some cases show how constraints on the given semantics carry over to the idiomatic one. But if someone wants to use the word *grammar* for semantic specifications too, I have no objections, although in this paper it is used in the technical sense of Hodges (2000).

Some idioms appear to have an atomic analysis, but for others, which apparently have practically the same distribution as their literal versions, an account along the lines of section 10 seems preferable. It allows the idiom to occur in familiar syntactic constructions, and computes its meaning with the familiar semantic operations. The crux, of course, is to provide the right atomic meaning 'chunks', and to prevent overgeneration.

This analysis is proposed in Nunberg et al. (1994), and an implementation is sketched in Sag and Wasow (1999, ch. 11). In this sense the analysis of section 10 is reminiscent of the HPSG analysis. But of course there are important differences. An HPSG grammar generates, in a sense, surface strings with appended feature structures containing all kinds of information: syntactic, semantic, morphological. In this paper I have adhered to a more traditional view of grammar, for example, in separating the semantics from the rest. As to overgeneration, HPSG can easily include in the lexical information about $pull^i$, say, that is must combine with $strings^i$ (in the right positions) and likewise for $strings^i$. This amounts to a version of the constraint (1.19) in the previous section which is both more precise and more general. And in this way all the examples of overgeneration concerning the idiom *pull strings* discussed here would be ungrammatical. But, to repeat, I have not entered into this sort of detail in this paper. And hopefully the discussion on an abstract level of some kinds of possible overgeneration has at least highlighted problems that a satisfactory account of idioms has to solve.

But aside from such details I believe there is one conclusion that may be safely drawn from the formal work in this paper: With respect to the principle of compositionality, idioms do not pose a particular problem.

References

Hodges, W. 1998. Compositionality is not the problem. *Logic and Logical Philosophy*, 6:7–33.

Hodges, W. 2000. Formal features of compositionality. *Journal of Logic, Language and Information.* To appear.

Janssen, T. 1997. Compositionality. In van Benthem, J. and A. ter Meulen, editors, *Handbook of Logic and Language*, pages 417–473. Elsevier, Amsterdam.

Nunberg, G., I. Sag, and T. Wasow. 1994. Idioms. *Language*, 70:491–538.

Sag, I. and T. Wasow. 1999. *Syntactic Theory: A Formal Introduction.* CLSI Publications, Stanford.

Westerståhl, D. 1999. Idioms and compositionality. In Gerbrandy, J., M. Marx, M. de Rijke, and Y. Venema, editors, *JFAK. Essays Dedicated to Johan van Benthem on the Occasion of his 50th Birthday.* Vossiuspers. Amsterdam University Press, on CD.