Same Syntax, Different Semantics: A Compositional Approach to Idiomaticity in Multi-word Expressions

Timm Lichte • Laura Kallmeyer

Abstract Idiomatic multi-word expressions (MWE) are commonly analyzed as phrasal units in syntax, in addition to their literal counterparts, and accordingly introduce syntactic rather than semantic ambiguity. However, an analysis of idiomaticity based on syntactic ambiguity is disadvantageous, because it neglects recent psycholinguistic findings about the processing of idiomatic MWEs, and it furthermore obscures the possible connection between their literal and idiomatic meaning. In this contribution, we sketch an alternative analysis, employing the framework of Lexicalized Tree Adjoining Grammar (LTAG), where idiomaticity is not subject to syntactic ambiguity, but emerges in the semantics.

Keywords multi-word expression · ambiguity · LTAG · Frame Semantics

T. Lichte, University of Düsseldorf, lichte@phil.hhu.de
L. Kallmeyer, University of Düsseldorf, kallmeyer@phil.hhu.de

In Christopher Piñón (ed.), Empirical Issues in Syntax and Semantics 11, III–140. Paris: CSSP.
http://www.cssp.cnrs.fr/eiss11/

1 Introduction

Multi-word expressions (MWE) consist of multiple lexemes that combine in some idiosyncratic, unpredictable or "idiomatic" way (Sag et al. 2002, Baldwin & Kim 2010). This combinatorial idiomaticity can manifest in different aspects of an MWE, for example in its syntax, semantics, pragmatics or statistics. In this work, we are mainly interested in semantic idiomaticity, which basically follows from the availability of a literal and an idiomatic meaning. Take, for example, the complex expression *to spill the beans*. Its literal meaning is 'to spill the beans' (obviously), whereas its idiomatic meaning is rather something like 'to divulge/reveal some secret(s)'.

Semantically idiomatic MWEs such as *to spill the beans* are said to be DECOMPOSABLE, whenever a quasi-bijective correspondence between com-

ponents of the literal and the idiomatic meaning can be established. The crucial pairs in the above example are 'spill'–'divulge/reveal' and 'beans'–'secret(s)'. However, this doesn't hold for all the semantically idiomatic MWEs, a typical example being *to kick the bucket* with its idiomatic meaning 'to die'. Those latter cases therefore get classified as NON-DECOMPOSABLE.

The ambiguity between literal and idiomatic readings can be modeled as the result of either syntax or semantics. In case of SYNTACTIC AMBIGUITY, the literal and idiomatic readings emerge from different syntactic derivations of the same sentence (see section 2). For example, there would be a literal *spill* and an idiomatic *spill* in the lexicon, or even more complex phrasal entries in the idiomatic case, that would independently and alternatively enter into the syntactic derivation.[1] On the other hand, SEMANTIC AMBIGUITY emerges when processing the lexical semantics of only one lexical entry for *spill*, therefore lacking phrasal entries altogether (see section 3). While earlier work, for the most part, has modeled MWEs in terms of syntactic ambiguity, we will elaborate the semantic ambiguity approach in this paper.

Semantic ambiguity approaches have a number of substantial advantages over syntactic ambiguity approaches: firstly, they seem to be more plausible in psycholinguistic terms, as there is evidence that the computation of literal and idiomatic meanings is based on the same syntax (Peterson & Burgess 1993, Wittenberg & Piñango 2011, Wittenberg et al. 2014); secondly, they simplify the parsing process, as only one syntactic derivation has to be performed and the disambiguation step can be delayed; finally, the connection between literal and idiomatic meanings can be made more explicit compared to syntactic ambiguity approaches where literal and idiomatic meanings are assigned to separate lexical entries. On the other hand, we will show that purported disadvantages of semantic ambiguity approaches dissolve under certain implementational assumptions, so that the advantages, particularly of lexical-semantic approaches, prevail.

We will base our implementation on the framework of Lexicalized Tree

[1]Note that by "syntactic derivation" we mean the derivational process, not just the derived syntactic structures.

Adjoining Grammar (LTAG). We conjecture, however, that the covered approaches can in principle be implemented in most other grammatical frameworks as well.

2 Idiomaticity as Syntactic Ambiguity

Syntactic ambiguity is triggered by semantically idiomatic MWEs whenever there are different syntactic derivations for literal and idiomatic meanings. This general approach is found in work across rather heterogeneous frameworks. While calling them "canonical form theories," Pulman (1993) mentions analyses from Transformational Grammar (Chomsky 1980), Lexical Functional Grammar (Bresnan 1982), and LTAG (Abeillé & Schabes 1989, Abeillé 1990, 1995). We might also add recent work in Head-driven Phrase Structure Grammar (Sailer 2003, Soehn 2006, Richter & Sailer 2009) and Sign-based Construction Grammar (Kay et al. in progress).[2] In the following, we will be focusing on LTAG.

2.1 LTAG

An LTAG (Joshi & Schabes 1997, Abeillé & Rambow 2000) consists of a set of ELEMENTARY TREES, which are lexicalized in the sense that at least one leaf node bears a lexical element, that is, a word token. These elementary trees can be combined to yield larger DERIVED TREES using either SUBSTITUTION or ADJUNCTION. Substitution is the replacement of a leaf node of a target tree with an elementary tree, whereas adjunction replaces a non-leaf, that is, an inner node with an elementary tree. Commonly, substitution is used in cases of complementation (including the subject), and adjunction in cases of modification.[3] An example is provided in figure 1.

LTAG is known for providing elegant accounts for a range of multi-word expressions with non-compositional meaning (e.g. Abeillé & Schabes 1996). The reason is that elementary trees of an LTAG can be made as large as is necessary to span any multi-word expression, even discontinuous or clausal ones, as elementary trees come with an extended domain of locality (EDL). This can also be observed in the example in figure 1. The EDL property is particularly useful when it comes to inflexible (*by*

[2] It seems that the approach of Fischer & Keil (1996) also runs into syntactic ambiguity.

[3] The exception to the rule are sentential complements, which usually combine with their governor via adjunction, in order to allow for long-distance extraction.

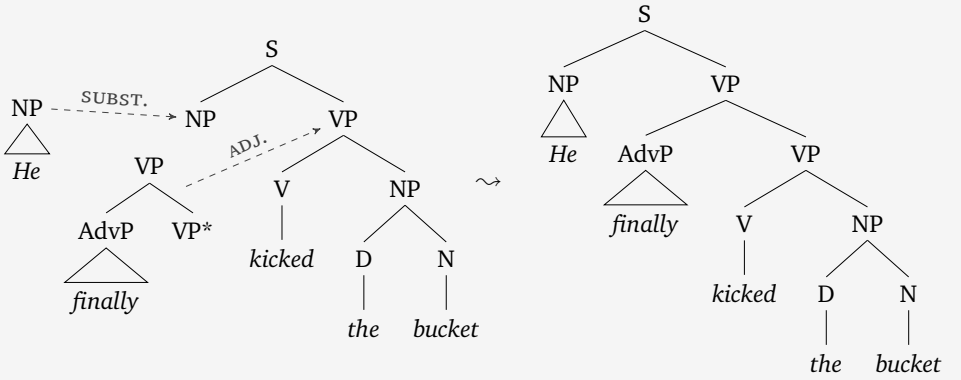

Figure 1 An LTAG derivation of *He finally kicked the bucket*

and large) or syntactically ill-formed MWEs (*kingdom come*), or MWEs with bounded words (*leave sb. in the lurch*). But also the greater flexibility of semantically idiomatic MWEs can be accounted for to some degree. An example is shown in the upper part of figure 2 with a frame-based semantics following Kallmeyer & Osswald (2013). Due to the flexible linking of syntax and semantics by means of interface variables (see the boxed numbers), internal and external modification can be adequately handled. In the idiomatic case, for example, the adnominal adjective *social* in *She kicked the social bucket* (meaning 'Socially speaking, she died') would adjoin to the N-node of *bucket*, but it would correctly modify the *dying* event thanks to the linking via $\boxed{0}$.[4]

By contrast, the literal reading of *kicked the bucket*, as can be seen from the lower part of figure 2, emerges from additional elementary trees in which *kicked* and *bucket* lexicalize separate elementary trees with some literal meaning representation. Hence, based on this sort of proliferation, there are two syntactic ways in which *kicked the bucket* can be derived.

2.2 Problems of Syntactic Ambiguity Approaches

While this sort of model clearly has its virtues, it nevertheless suffers from the disadvantages of syntactic ambiguity approaches already mentioned in section 1, which will be elaborated in the following.

One crucial peculiarity of the model just presented is that it enumerates

[4]See a similar approach using Synchronous TAG in Sailer 2003:(438).

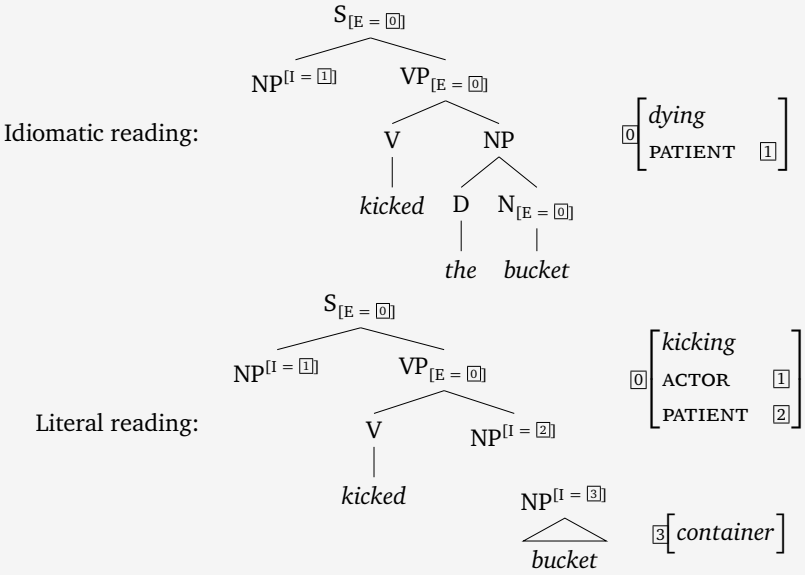

Figure 2 Syntactic ambiguity of *kicked the bucket* induced by disjunction over pairs of LTAG elementary trees and frame semantic representations

idiomatic and literal expressions by assigning a separate elementary tree (or a family of trees) to each of them. Hence, both meanings are based on purely homonymous words, or to put it differently, it falls out as a mere coincidence that words such as the literal *kick* and the idiomatic *kick* happen to be pronounced similarly. This has two immediate consequences: (i) the possible connection between literal and idiomatic meanings of an MWE, for example in terms of register or etymology, is obscured; (ii) regarding the flexibility of idiomatic MWEs, generalizations about variability are at risk of being missed, as Pulman (1993:256–257) notes. He gives examples of lexical variability such as in *put/lay/spread the cards on the table*, and of constructional variability as in *let the cat out of the bag* versus *the cat is out of the bag*. Looking at elementary trees alone, it is obvious that both drawbacks cannot easily be argued away. However, elementary trees of an LTAG are usually described in a metagrammar (using, e.g., XMG; Crabbé et al. 2013), which helps to express generalizations across elementary trees. Hence, within such a metagrammar, it seems rather straightforward to capture at least the lexical and constructional variability that

Pulman brings up.

A more severe drawback touches upon the predictions made regarding the processing costs of MWEs: a syntactic ambiguity approach predicts that MWEs are syntactically more demanding during processing, since potentially two derivations have to be computed, one for the literal and one for the idiomatic meaning. However, psycholinguistic findings seem to suggest that processing costs emerge in the semantics rather than in syntax (Peterson & Burgess 1993, Wittenberg & Piñango 2011, Wittenberg et al. 2014). This contradicting evidence is fostered by general considerations about computational economy: given a parser that implements a syntactic ambiguity approach, it would act inefficiently when parsing MWEs, since it would create the very same syntactic structure several times. Therefore, speaking of parsing efficiency, one would clearly like to place ambiguity at the level of semantics in order to delay disambiguation.

Next, there are a couple of more linguistic problems when accounting for partial uses of decomposable MWEs, that is, instances of pronominalization or isolation of NP-components of an MWE, let alone the "extendability" (Egan 2008) of literal MWE senses. An example of pronominalization due to Riehemann (2001:(229)) is given in (1):

(1) Eventually she spilled all the beans. But it took her a few days to spill them all.

The critical part is *spill them* in the second sentence, where the canonical NP-component of the MWE *spill beans* is replaced by a pronoun.[5] Within an LTAG approach, one could pursue one of at least two modeling strategies: (i) treat the pronoun as a lexicalized leaf node similarly to the NP-component in *spill beans*, or (ii) treat the NP-component as an unlexicalized leaf node into which only *beans* and *them* can be substituted, hence assigning it the status of a highly restricted argument slot. The first strategy would be liable to lose contact with a general theory of pronominalization,[6] while the second strategy would make some ad hoc categories

[5]Riehemann (2001:207) tentatively assumes that pronominalization in *spill them* is licensed by a more general, metaphorical use of *spill*, which allows for combinations with a wider range of NPs, such as in *spill the secrets*.

[6]Thanks to Manfred Sailer for pointing this out.

necessary in order to rule out lexical anchors other than *beans*.

A case comparable to pronominalization, also having to do with the flexible use of MWEs, is the isolation of the NP-component. The following example is due to Manfred Sailer and Sascha Bargmann (pers. comm.):

(2) Pat pulled some strings for Chris. But Alex didn't have access to any strings.

In the second sentence, the NP-component *any strings* occurs isolated from the rest of the MWE, namely, the verb *pull*, while still bearing the idiomatic meaning 'connections'. This sort of isolation can only be modeled by a special elementary tree of *strings*, which is, however, difficult to limit to certain discourse contexts as in (2).

Maybe an even more extreme case of flexibility is observed by Egan (2008:(13b)):

(3) If you let this cat out of the bag, a lot of people are going to get scratched.

The remarkable property of the conditional in (3) is that *to get scratched* resumes the figurative mapping of *let this cat out of the bag*, even though *let this cat out of the bag* is part of the antecedent clause. Egan (2008) therefore calls *to get scratched* an EXTENSION of the MWE and he rightly suspects that syntactic ambiguity approaches face difficulties in covering these extensions. The only viable strategy, so it seems, is to extend the grammar by (probably masses of) secondary MWEs such as *to get scratched*, which would still be hard to limit to contexts which contain the primary MWE.

The last group of intricate uses of idiomatic MWEs that we will address here is related to modification. In the previous section, we have already seen that simple cases of internal and external modification can be dealt with using the appropriate linking pattern between syntactic nodes and positions in the frame semantic representation. This only works for modifiers pertaining to the idiomatic sense of the MWE. However, as Ernst (1981:(27)) has shown, the modifier can add to the literal meaning as well:

(4) The federal agency decided to take the project under its well-muscled

wing.

Here *well-muscled* is supposed to mean something like 'powerful'. Ernst also provides an example of idiomatic adnominal modification with *federally-funded*. In fact, it is not difficult to come up with an example in which these sorts of modification, which pertain to either idiomatic or literal meaning dimensions, are mixed:

(5) The federal agency decided to take the project under its well-muscled, federally-funded wing.

A similar case of instant multi-dimensionality is found in Ernst's category of "conjunction modification," which he exemplifies with the following datum (Ernst 1981:(10)):

(6) Malvolio deserves almost everything he gets, but ... there is that little stab of shame we feel at the end for having had such fun pulling his cross-gartered leg for so long.

The adjective *cross-gartered* modifies the literal meaning of *leg*, which "refers to Malvolio's real, flesh-and-blood leg," but it is not figuratively mapped onto the co-existing idiomatic meaning of *pull sb.'s leg*. Ernst resolves this as an extra proposition that is added through conjunction – therefore the name. It is not at all obvious how literal, mixed, and conjunction modification could be satisfactorily treated within a syntactic ambiguity approach. Since the connection between literal and idiomatic meaning is generally obscured, the modifiers, like the MWEs, would have to be multiplied for each of the meaning dimensions.

To avoid false expectations: we won't solve all these flexibility issues here. This would eventually mean to close the gap to a profound theory of metaphor. Still, semantic ambiguity approaches, particularly the inference-based approaches, seem to be a better starting point.

3 Idiomaticity as Semantic Ambiguity

Put simply, semantic ambiguity emerges whenever the literal and the idiomatic reading of an MWE cannot be traced back to syntactic ambiguity. Hence there is only one syntactic derivation for both readings, and

the semantic ambiguity must be either induced by lexical specification or by global quasi-inference rules. Compared to syntactic ambiguity approaches, semantic ambiguity approaches are applied rather seldom, let alone in a formally explicit fashion. In the next two sections we will first review earlier work, and then present our own implementation based on LTAG and frames in section 4.

3.1 Previous Lexical-semantic Approaches

The only formally more or less explicit lexical-semantic approach that we are aware of has been presented by Gazdar et al. (1985:sect. 10.7). In their proposal, the components of decomposable MWEs are assigned two meaning constants (of Intensional Logic), as can be seen from (7a), namely, one for their literal and one for their idiomatic meaning. We separate them with the ambiguity symbol || from Wurm & Lichte (2016):

(7) a. *spill* := spill′ || spill-idiom′
beans := beans′ || beans-idiom′
b. spill-idiom′ (beans-idiom′): defined
spill-idiom′ (beans′): undefined
spill′ (beans-idiom′): undefined

These meaning constants are interpreted as partial functions (contrary to what was usual in Intensional Logic at that time).[7] The reason to choose partial functions is that this makes it possible to restrict the emergence of idiomatic meanings to the complete occurrence of the MWE. Thus, literal and idiomatic meaning components cannot be properly combined, since the result would be undefined such as in (7b).

Note that Gazdar et al. (1985:244) propose to treat non-decomposable MWEs as "syntactically complex lexical items," hence within a syntactic ambiguity approach. However, we think that in principle partial functions can also be used when dealing with non-decomposable MWEs:

(8) a. *kick* := kick′ || kick-idiom′
bucket := bucket′ || bucket-idiom′
b. kick-idiom′ (bucket-idiom′): defined
kick-idiom′ (bucket′): undefined

[7] Pulman (1993) therefore calls it a "partial-function approach."

kick′ (bucket-idiom′): undefined

As with decomposable MWEs, the interpretation is only defined if the right meaning constants are combined via functional application. Of course it needs to be clarified what bucket-idiom′ denotes. In section 4 we will basically state that it has the same denotation as kick-idiom′, namely, 'die', and both denotations get identified upon composing kick-idiom′ and bucket-idiom′.

The lexical-semantic approach of Gazdar et al. has several general advantages over the syntactic ambiguity approach based on LTAG: it yields a unified syntax for idiomatic and literal readings, and, following this, appears to be psycholinguistically more realistic. However, it also comes with considerable, general drawbacks. One is the invention of masses of meaning constants that essentially reflect morphological properties. There is no genuinely semantic motivation for having something like a spill-idiom′ predicate, when it conceptually coincides with divulge′. These predicates are needed only in order to capture constraints on the surface structure, that is, at word level.

Another drawback is computational, as Pulman (1993) points out, namely, the introduction of extra ambiguity and following this a "considerable combinatorial explosion." This might come as a surprise given that yielding a unified syntax was thought to delimit computational effort. The reason is that ambiguity resolution now takes place at word level, not at the phrasal level. Therefore, the grammar has to try out many illicit combinations of idiomatic and literal word meanings, without taking phrasal information into account.

Pulman, furthermore, claims that Gazdar et al.'s approach either under- or overgenerates, for example, when treating the following relative clause (Pulman 1993:(50)):

(9) He tried to brake the ice which inhibited our conversation.

The issue arises when the relative pronoun *which* receives the idiomatic interpretation of *ice*. If this is the case, then the interpretation of the verb of the relative clause, *inhibited*, must be made compatible, that is, ambiguous. However, this then also licenses the idiomatic interpretation of sentences like the following (Pulman 1993:(51)):

(10) The ice inhibited our conversation.

Pulman questions the immediate availability of the idiomatic interpretation of (10) in contrast to (9).

Similarly, the partial-function approach is at risk of overgeneration when dealing with cases of partial use and, in particular, extendability such as in (3) – at least it will add substantially more ambiguity. Regarding literal, mixed and conjunctive modification, the situation is even worse: it is not conceivable how this approach, in which idiomatic and literal meaning dimensions remain de facto disconnected, could satisfactorily handle those cases. What Gazdars et al.'s partial-function approach can rather nicely deal with, however, are cases of lexical variability, since variants can be assigned the same idiomatic meaning constant (Gazdar et al. 1985:239–240).

Lastly, note that the relation between distinct meaning potentials is notoriously unclear in lexical semantics. In (7) and (8), we used the symbol || to discriminate between literal and idiomatic meaning, borrowing it from Wurm & Lichte (2016). But what does || mean or correspond to? The obvious choice, namely, disjunction, is far from adequate (see Poesio 1996, Wurm & Lichte 2016): The propositional meaning of *If he kicked the bucket, the water would spill over the floor* is not something like ('he died' $\vee$ 'he kicked the bucket') $\rightarrow$ 'the water would spill over the floor', or the equivalent ('he died' $\rightarrow$ 'the water would spill over the floor') $\wedge$ ('he kicked the bucket' $\rightarrow$ 'the water would spill over the floor'). What is the relation then? And how does it work out compositionally? As far as we can see, Gazdar et al. remain silent about these fundamental questions. See Wurm & Lichte 2016 for some general algebraic considerations.

3.2 Previous Inference-based Approaches

In the light of the problems encountered in the lexical-semantic approach of Gazdar et al. (1985), Pulman (1993) proposes to deduce the idiomatic meaning from the literal one by means of "quasi-inference."[8] In this approach, MWE-components are equipped with their literal meaning only, whereas the idiomatic meaning comes in later once the complete MWE has been seen. An example of the style of such quasi-inference rules is

[8]Another, yet very informal, inference-based approach is found in Egan (2008).

shown for the MWEs *kick the bucket* and *spill the beans* in (11a):

(11) a. $\forall x, y.\text{kick}'(x, y) \wedge \text{bucket}'(y) \approx \text{die}'(x)$
b. $\forall x, y.\text{spill}'(x, y) \wedge \text{bean}'(y) \approx$
$\exists z.\text{divulge}'(x, z) \wedge \text{information}'(z)$

While it is obvious that (11a) and (11b) are not meant to be regular inference rules (since the left-hand side can be false with the right-hand side being true, and vice versa), Pulman remains vague about the exact meaning of $\approx$. It is supposed to mean that if the left-hand side is "matched" in the logical form, the right hand side meaning is "possible" as well, "perhaps" replacing the literal meaning "depending on the context" (p. 262). Furthermore, as Pulman's inference rules are purely semantic, he needs to limit their scope by "lexical indexing," that is, by attaching a set of lexical items to each of the rules. For (11a), this would be the set {*kick*, *the*, *bucket*}, which would have to be a subset of a given sentence in order for the rule to apply.

Similarly to Gazdar et al.'s partial-function approach, Pulman's quasi-inference rules treat the distinction between literalness and idiomaticity mainly in the semantics, which leaves the syntax unified and thereby consistent with psycholinguistic findings. In contrast to partial functions, however, quasi-inference rules seem to reduce the degree of local ambiguity, because the idiomatic meaning does not emerge per word. Instead, it is based on larger, that is, phrasal chunks of literal meaning. Moreover, the domino effect of artificial idiomatization, which Pulman showed with the relative clause in (9), can be avoided (let alone partial functions in general).

The main problem of Pulman's implementation is its vagueness, which makes it virtually impossible to see how extendability, challenging cases of modification, etc. can be treated without running into vast overgeneration or yielding an incorrect truth-conditional semantics. Even in basic cases, the use of lexical indexing for restricting quasi-inference seems to be too permissive.[9] It is also not settled that the inference-based approach

[9]There is one aspect of overgeneration, however, that is deliberately prompted: quasi-inference rules do not check for the syntactic construction from which the literal meaning emerges. Therefore, they can be applied even to passive constructions like *the bucket*

is indeed computationally lighter than the lexical-semantic approach, for quasi-inference rules seem to be potentially non-monotonic and might apply at any time and in any order. Finally, it should be borne in mind that the sharp procedural distinction between literal and idiomatic meaning is not uncontroversial in the psycholinguistic and philosophical literature (see, e.g., Récanati 1995, Gibbs 2002, Wearing 2012).

4 A New Lexical-semantic Approach

The main problem of the lexical-semantic approach of Gazdar et al. (1985) is that it fuses morphological constraints with semantic representations. What we therefore propose is to disentangle these two aspects and treat semantics and morphology as separate but interrelated dimensions. To this end, we enrich the frame semantic representations from figure 2 with SEM and MORPH features, while the syntax remains a regular LTAG. Because special elementary trees for MWEs, such as the first one in figure 2, are missing, the morphological features are necessary for confining the context where the idiomatic meaning emerges.

4.1 Elementary Structures

Our proposal is fleshed out in figure 3 based on the lexical entries for *kicked* and *bucket*. In either case, the first ||-disjunct corresponds to the literal meaning and the second one to the idiomatic meaning. Similarly to the syntactic ambiguity approach in figure 2, the elementary trees and the SEM-MORPH representations are linked via interface variables (see the boxed numbers). As for *kicked*, the subject NP (with variable 1) is linked to the actor of a *kicking* frame in the literal case, and to the patient of a *dying* frame in the idiomatic case.[10] The object NP (with variable 2), however, is linked to some component of the verbal frame only in the literal case, whereas, in the idiomatic case, it is linked to the verbal frame as

is kicked or *the breeze was shot* and yield their idiomatic meaning, contrary to what is consensus in the literature (Sag et al. 2002, Baldwin & Kim 2010). Pulman argues that those constructions are incompatible with idiomatic interpretations on pragmatic, that is, information-structural grounds. We will largely ignore aspects of constructional fixedness in this work.

[10] Boxed numbers with a prime are a proxy for a link to the SEM part, more precisely, they invoke a path equation of the following kind: $\boxed{n}' = \boxed{n}\text{SEM}$. This means that $\boxed{n}'$ is the value of SEM of $\boxed{n}$.

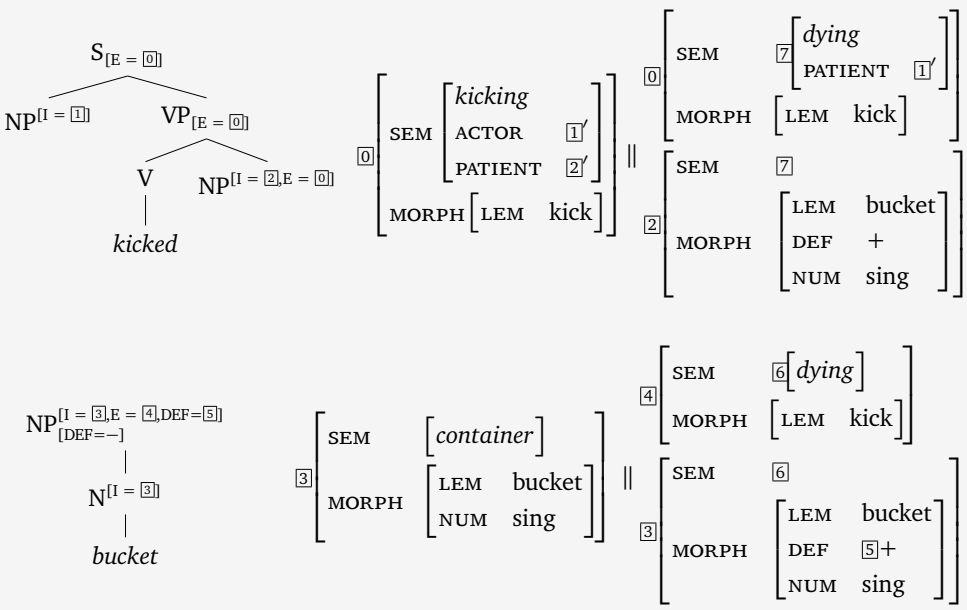

Figure 3 Entries for *kicked* and *bucket* within a semantic ambiguity approach based on LTAG

a whole, while the MORPH feature anticipates the substitution of *bucket*. More precisely, the object NP points to a SEM-MORPH structure in which the SEM value is identical with the SEM value of the SEM-MORPH structure of the verbal projection (with variable 0).[11] Yet as different as they may be, the literal and idiomatic ||-disjuncts are explicitly connected by sharing interface variables, namely, 0 and 1′, which link them to the same elementary tree.

One noticeable property of this proposal, which we call IDIOM MIRRORING, is that the idiomatic meaning is spread over all the components of the MWE. Therefore, also the NP-component carries the meaning of the whole, that is, *bucket* carries the *dying* frame in figure 3. While this might look odd at first, it is necessary in order to allow modification at *bucket* (via adjunction at the N-node, see section 4.3) to yield wide scope. Idiom mirroring is justified on independent grounds as well, since *bucket*,

[11] Note that the value of MORPH is a simple feature list rather than a recursive typed feature structure, which is the value of SEM following Kallmeyer & Osswald (2013).

in some cases, can contribute the idiomatic meaning even when isolated from *kicked*, for example, in *bucket list*.[12]

Note that the determiner *the* does not take part in idiom mirroring in figure 3. This seems justified, as *the* is a semantically bleached function word. Instead the adjunction of standard *the*, which aims at the NP-node of *bucket*, is enforced by the otherwise pending feature clash of the DEF features.[13] It should be stressed that the features in figure 3 (and the figures to come) are chosen for expository reasons. It would be equally viable to use DEF only on the morpho-semantic side, and express the fact that the count noun *bucket* generally requires a determiner by means of another more generic feature (e.g., DET). Conversely, to rule out definite determiners that do not yield the idiomatic reading as in *He kicked his bucket*, the DEF feature could be further refined. A far more intricate question is what happens to the semantic contribution of *the*, bleached as it may be, in the idiomatic case.[14] Assuming that *the* contributes some sort of definiteness operator that selects a specific antecedent from context, the linking in figure 3 would predict that, upon adjoining *the* into the NP-node of *bucket*, the definiteness operator will take wide scope over the dying event. Thus, *kicking the bucket* would denote a specific dying event. This, however, seems questionable given embeddings such as *Kicking the bucket was easy* that clearly lack such a denotation. Yet, working out the subtle details of this part of the story must be left to future work.

It is instructive to compare this with the analysis of a decomposable MWE such as *spill beans*. The relevant entries are shown in figure 4. Comparing them to the entries in figure 3, the high degree of structural similarity is striking. Firstly, the syntactic trees are basically the same, except for the lexical anchor of course, and they conform to the commonly assumed elementary trees in LTAG. Secondly, once again we make use of idiom mirroring in the idiomatic meaning components, even though the

[12] Thanks to Manfred Sailer for pointing out this fact.

[13] In LTAG with feature structures, every node consists of top and bottom feature structures that get eventually unified, that is, after substitution and adjunction have taken place. Therefore, if no determiner was adjoined at the NP-node of *bucket*, the equation $\boxed{5} = -$ would hold after top-bottom unification, which contradicts the specification $\boxed{5} = +$ on the idiomatic side.

[14] Thanks to Jamie Findlay for making us aware of this issue.

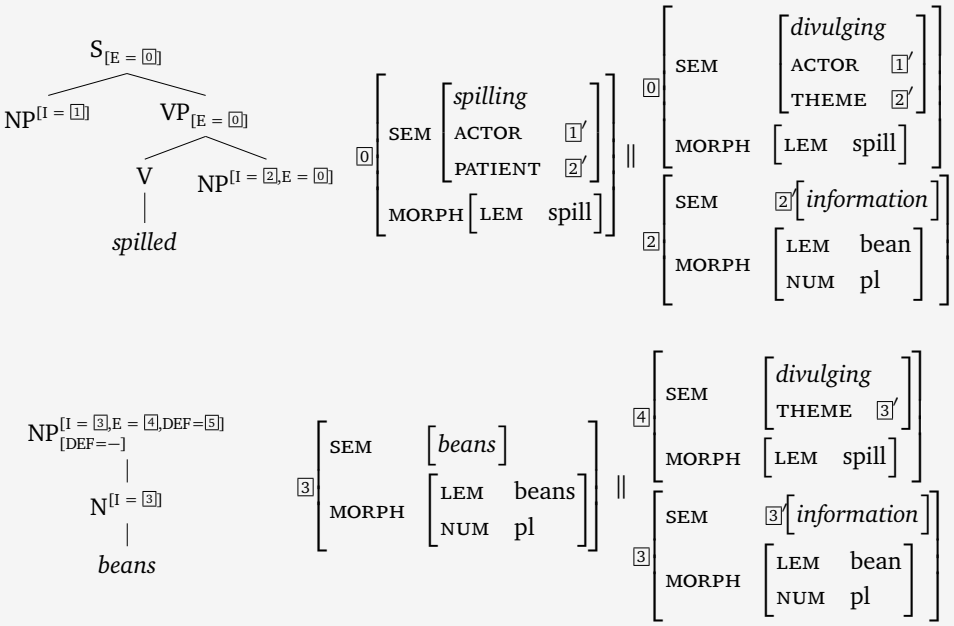

Figure 4 Entries for *spilled* and *beans* within a semantic ambiguity approach based on LTAG

SEM values of 3 and 4 now differ in order to allow for internal modification (see section 4.3). This sort of uniformity is particularly rewarding from the point of view of the metagrammar (see section 2.2), because it supports a lean description of the generalizations across those and other entries.

4.2 Composition with ||-Disjunctions

Following Kallmeyer & Osswald (2013), upon substituting or adjoining elementary trees, the feature structures of affected nodes are unified, and consequently their interface variables. Thus when *bucket* gets substituted into the object NP leaf of *kicked*, the identities $2 = 3$ and $0 = 4$ are obtained. Furthermore, the ||-disjunctions are unified in a straightforward distributional way by which every ||-disjunct gets unified with each of the

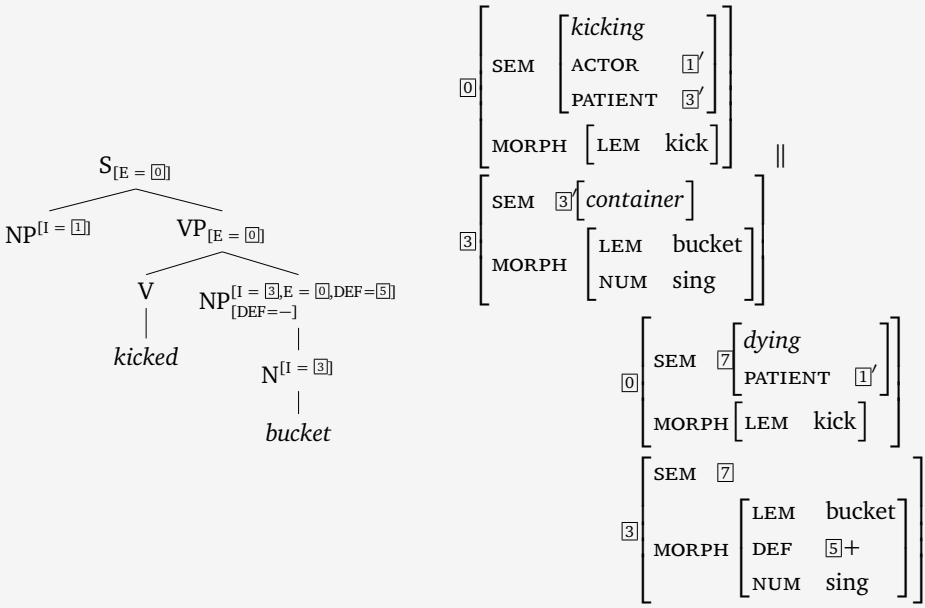

Figure 5 Result of substituting the elementary tree of *bucket* into the elementary tree of *kicked* (the determiner *the* still needs to be adjoined)

||-disjuncts of the other tree (but only two times successfully).[15] The result of all this is shown in figure 5. Note that the SEM-MORPH side remains multi-rooted, which is intended, and connected to the syntactic tree.

There are two other important ramifications of this compositional approach: (i) the frame of the verbal head must be visible in the NP-slot (via the E feature), otherwise the idiomatic meaning of *bucket* would be generally available; (ii) the approach seems to suggest that idiomatic *kick the bucket* is syntactically as flexible as the literal counterpart, which many would consider too permissive. In fact, we adopt the view of Pulman (1993) that syntactic inflexibility results from a mismatch between information structure and semantics.[16] But if this view was not shared, one could flexibly adjust the ||-disjuncts of, for example, passive elementary

[15] In formal terms, the following distributional equation holds: $(a \parallel b) \sqcup (c \parallel d) = (a \sqcup c) \parallel (a \sqcup d) \parallel (b \sqcup c) \parallel (b \sqcup d)$

[16] In fact, Manfred Sailer has informed us that there is (rare) evidence for passivization of idiomatic *kick the bucket*, namely, *The bucket will be kicked*.

trees to only ship literal meanings.

Finally, the readings of a lexeme are immediately available when instantiating its elementary tree. This could be taken to contravene psycholinguistic results that suggest that readings are not equally accessible. Putting distinct weights on the disjuncts, however, might solve this issue, and ease the potential of combinatorial explosion.

Hence, when comparing this implementation of a semantic ambiguity approach with the one of Gazdar et al. (1985) sketched in section 3.1, two substantial improvements become evident: firstly, there are no ad hoc meaning constants, and morphosyntactic restrictions are expressed where they belong; secondly, local ambiguity is considerably diminished, since phrasal cues can be used straightforwardly. But also the empirical coverage can be improved on, as the following section will show, where we will treat cases that were considered challenging in the preceding sections.

4.3 Analysis of Modification and Partial Use

Recall the sorts of adjectival noun modification inside idiomatic multiword expressions we mentioned in section 2:

(12) a. He kicked the political bucket. (external mod.)
'Politically speaking, he died./His political life ended.'

b. He spilled the political beans. (internal mod.)
'He revealed political secrets.'

c. We took it under our well-muscled, federally-funded wing. (mixed mod.)
'We strongly protected it with the aid of federal funds.'

d. We pulled his cross-gartered leg. (conjunctive mod.)
'We teased him and he had a cross-gartered leg.'

External and internal modification in (12a) and (12b) are treated similarly to the syntactic ambiguity approach, namely, by adjunction at the N-node of *bucket* and *beans* respectively. The different scopes of the modifiers result from the different linking by means of interface variables: in the case of external modification, the N-node is linked with the idiomatic meaning of the whole MWE (in figure 3 with 'die' via 3), whereas in the case of internal modification, there is a link with the proper idiomatic meaning of the noun (in figure 4 with 'information' via 3). A tentative entry for the

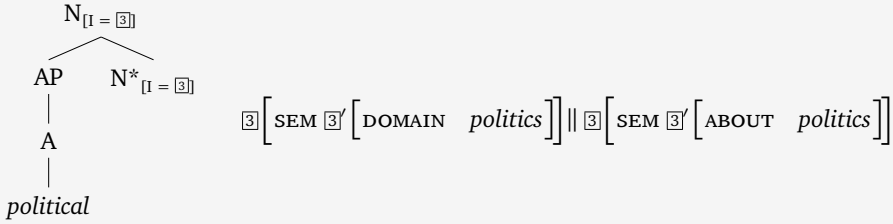

Figure 6 Entry for the adjectival nominal modifier *political*

adjective *political*, which can be used for both sorts of modification and is therefore ambiguous, is shown in figure 6.[17] The first ||-disjunct is supposed to cover the external modification case and is therefore necessarily vague, as it is hard to pin down exactly what the meaning of those "domain delimiters" (Ernst 1981) is.[18] The resulting derived tree after adjoining the entry for *political* into the derived tree of *kick bucket* from figure 5 is shown in figure 7. Note that the semantic composition arises from the final top-bottom unification at the N-node that dominates *political bucket*. In figure 7, this unification has already been performed. In this example, we assume that the ABOUT feature in the second ||-disjunct of *political* is incompatible with both types, *container* and *dying*. Therefore, only the DOMAIN reading shows up.

What makes (12a) also challenging, is that there seem to be two figurative interpretations in a row: at first, *kick the bucket* is interpreted as 'die' or 'ending of a (biological) life', and upon adding *political*, the meaning 'ending of a political career' emerges, drawing upon the general conceptual metaphor that 'career' can be seen as 'life'. Moreover, note that some features seem to be prohibited, for example MANNER: one cannot use *painful* as a modifier of *bucket* with the idiomatic meaning that the manner of dying was painful:

(13) #He kicked the painful bucket.
'He died painfully.'

[17] The presented adjectival entries are tentative in the sense that they are not at all meant to be exhaustive, but to only cover some distinctions that are immediately relevant to make our examples work.

[18] Metalinguistic modifiers such as *proverbial* are similar in this respect.

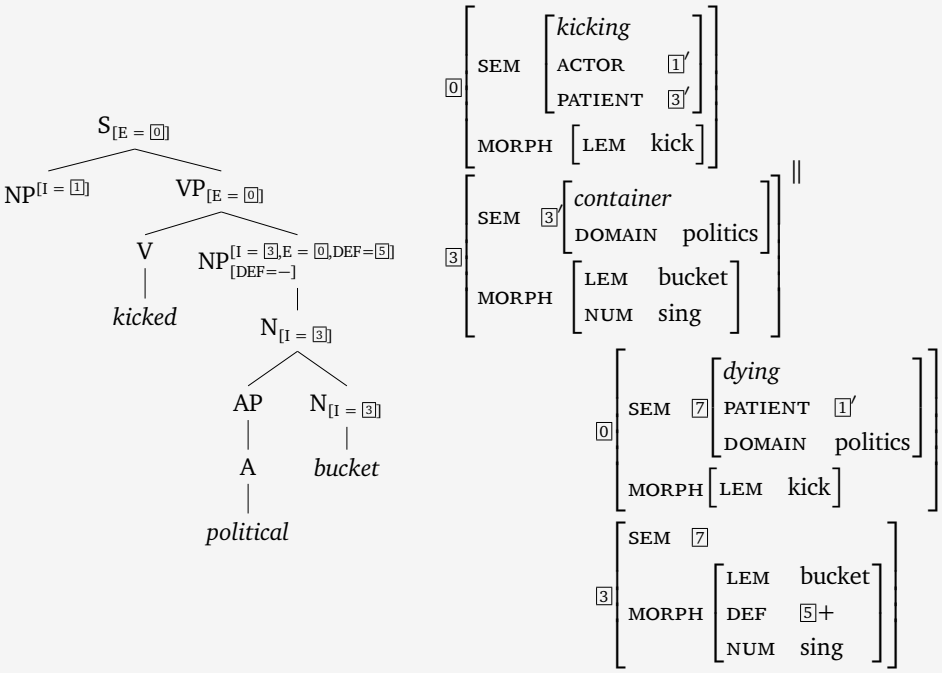

Figure 7 Result of adding the lexical entry in figure 6 to the derived structure in figure 5; the tree of *political* has been adjoined to the N-node of *bucket*

Surprisingly enough, the option with adverbial modification is acceptable on an idiomatic reading:[19]

(14) He painfully kicked the bucket.
'He died painfully.'

Given these observations, it seems possible to modify the manner denotation of the MWE from the "outside" – but why not from the "inside"? One explanation could be that something is wrong with our analysis, for example, the linking of the 1 feature and the overall event (see label 3 in figure 7), or the assumption that *bucket* mirrors the idiomatic meaning of the whole MWE. Fortunately, there is another and, in our opinion, far more interesting explanation, namely, that the idiomatic interpretation

[19] We owe Christopher Piñón this observation.

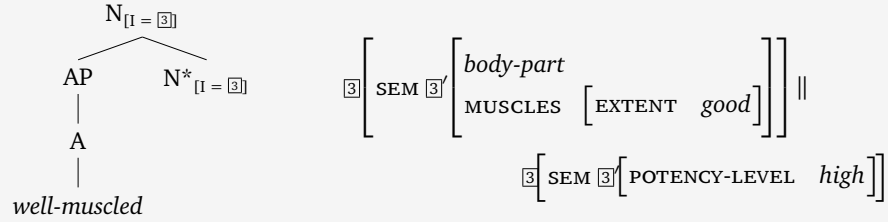

Figure 8 Entry for the adjectival nominal modifier *well-muscled*

of *kick the painful bucket* is unavailable because of the awkward *literal* interpretation of *painful bucket*. This might look like an ad hoc escape hatch, but, in fact, the persistence of literal fragments under an overall idiomatic interpretation can also be observed with conjunctive modification (see below). Therefore, we think that this line of thought should be taken seriously, even though the ramifications within our framework are hard to assess at this point. Given these complications, we leave it to future work to examine whether the general ambiguity approach sketched in figure 6 is actually sustainable.

The challenge of mixed modifications such as in (12c) is that one of the modifiers (*federally-funded*) pertains to the idiomatic meaning of the MWE, while the other one (*well-muscled*) modifies the literal part. Thus, there are rather two questions: (i) how does *well-muscled* access the literal meaning of the MWE, and (ii) how is this transferred to the idiomatic meaning, to which *federally-funded* applies? Note that the linear order of *well-muscled* and *federally-funded* is not fixed. The simplest and somewhat obvious solution is to make use of ambiguous entries again.[20] This strategy is followed in the entry in figure 8, where the first ||-disjunct corresponds to the literal meaning, and the second one to the idiomatic meaning.[21] Now, the question is: how can this possibly *not* overgenerate? What prevents the combination of the literal part of *well-muscled* with the idiomatic

[20] Another more ad hoc solution would be to encode the idiomatic part of the modifier already in the entry of the MWE. However, since the class of adjectives that can modify idiomatic *wing* is presumably not closed, and since *well-muscled* is amenable to this interpretation also with other nouns, this solution is not preferred.

[21] Note that the idiomatic meaning of *well-muscled* is not stipulated in accordance with the idiomatic meaning of *wing* but is seen to follow from general conceptual metaphors that can be formalized as quasi-inference rules (see section 5).

part of *wing*? The answer is: the type system. Remember that the frame semantic objects are typed, hence we assume that the literal type (*body-part*) is not unifiable with the idiomatic type of *wing*. Therefore, what will happen is that only type-compatible parts will unify, regardless of whether they correspond to the literal or idiomatic meaning.

However, this ambiguous-entries approach comes to its limits when conjunctive modification as in (12d) is considered. The crucial difference from the cases of the last two paragraphs consists of the peculiar way the modifier gets interpreted: the adjectival modifier *cross-gartered* is interpreted only literally, while the MWE *pull his leg* may nevertheless receive an idiomatic interpretation. In other words, the modifier may drop out of the idiomatic interpretation, still (or for this very reason) giving rise to an inference based on the literal interpretation. A possible interpretation of (12d) therefore is 'we made fun of him and he has a cross-gartered leg'. But how can we yield this interpretation with an approach like in figures 6 and 8 that separates literal and idiomatic meaning into different ||-disjuncts? The prospects seem to be the following, unfortunately: either the idiomatic interpretation does not emerge, because it is incompatible with the literal meaning of *cross-gartered*, or we add to *cross-gartered* some compatible, yet bleached idiomatic meaning. In the latter case, which is problematic in many respects, the "conjunctive" proposition ('he has a cross-gartered leg') could nevertheless get lost, since there would be at least one ||-disjunct with the bleached idiomatic meaning that could be used by itself. The basic problem therefore seems to be that there is no way to keep track of which ||-disjunct represents the idiomatic meaning, and which one the literal meaning. If this distinction was available, the condition could be imposed that the bleached idiomatic meaning may be used only if the literal meaning applies too. In this respect, the inference-based approach, to which we will turn in section 5, seems to be better off, as it sharply distinguishes between those two. However, we think that the inference-based approach eventually runs into the same difficulties as the lexical-semantic approach. What is needed instead is the possibility for the propositional interpretation of *parts* of the literal meaning, for example, the literal interpretation of *cross-gartered leg*, but not of *pull*, in (12d). It is not yet clear to us whether this kind of granularity can be achieved within the presented ambiguity framework.

Another challenging sort of modification that we mentioned in section 3 is modification by a relative clause as in (15), repeated from (9):

(15) He tried to brake the ice which inhibited our conversation.

Pulman (1993) argues that this sort of modification is problematic for Gazdar et al.'s lexical-semantic approach: since *which* is assigned the idiomatic meaning of *ice*, one has to assume a compatible partial function to be the meaning of *inhibited*. This, however, leads to overgeneration as *inhibited* now can combine with idiomatic *ice* alone. Fortunately, this issue does not arise within the presented LTAG approach, simply because of what we have called idiomatic mirroring above: idiomatic *break* not only constrains the object noun to be *ice* (similarly to Gazdar et al.'s partial functions), but *ice* as well constrains the governing verb to be *break*. Hence the constraints are effective in both directions, and this ultimately prevents idiomatic *ice* from going astray. Moreover, thanks to the division between SEM and MORPH, the relative pronoun *which* can be made to only refer to the semantics of idiomatic *ice*, so that *inhibited* may remain agnostic concerning the idiomatic/literal status of the semantics of the relative pronoun.

Finally, in section 2, we discussed three sorts of a partial use of MWE-components, repeated in (16) (excerpts from (1), (2) and (3)):

(16) a. [spill beans] ... to spill them all. (pronominalization)
b. [pull strings] ... didn't have access to any strings. (isolation)
c. [cat out of the bag] ... a lot of people are going to get scratched. (extension)

The common feature of (16a)–(16c) is that there is a close, preceding discourse context in which the full MWE is overtly realized. This specificity of the context is indicated by the material inside squared brackets. Hence, pronominalization of *beans* in (16a) is said to be only possible if there is a full realization of *spill beans* in the preceding context, and similarly in (16b) and (16c). To capture this in an adequate way, an approach would have to allow for the access of discourse information in defining, for example, the anaphoric function of *them* and relaxing the constraints of idiomatic *spill* so that it can combine with *them* given an appropriate con-

text. It is obvious that our approach, as is, cannot account for partial use, for the simple reason that discourse structure is not part of the model. It is also obvious, however, that we will need concurrent access to literal and idiomatic meaning (particularly for the case of extension in (16c)), and that we therefore have to look for the best-fitting model among the semantic-ambiguity approaches, not the syntactic-ambiguity approaches.

In sum, the presented lexical-semantic approach based on LTAG and SEM-MORPH descriptions supports a unified, compositional syntax, avoiding some technical and empirical shortcomings of the partial-function approach of Gazdar et al. (1985). We also showed that our approach can handle a range of challenging cases of modification, including cases of modification by relative clauses that are considered problematic for Gazdar et al.'s approach. Yet it remains to be seen whether and how cases of partial use can be integrated, once discourse structure is available.

5 A Similar Inference-based Approach?

In section 3.2, we discussed Pulman's proposal of quasi-inference rules, that is, global entailments, as an alternative to our and Gazdar et al.'s lexical-semantic approach. Those entailments are global in the sense that they in principle apply independently from both the lexicon and the syntactic derivation. We concluded that the formalization of quasi-inference rules, at least as far as Pulman's implementation is concerned, leaves much to be desired. In this section, we try to explicate the notion of quasi-inference rules by targeting some of the central intuitions that Pulman expresses.

Recall that Pulman would probably write down the following quasi-inference rule to deduce the idiomatic meaning of *kick the bucket* from the literal one (repeated from (11a)):

(17) $\forall x, y.\text{kick}'(x, y) \wedge \text{bucket}'(y) \approx \text{die}'(x)$

The big question here is what $\approx$ is supposed to mean. Pulman (1993:262) paraphrases it in the following way: if the left-hand side is "matched" in the logical form, the right-hand side meaning is "possible" as well, "perhaps" replacing the literal meaning "depending on the context." Hence, it is clear is that quasi-inference is not to be confused with regular inference.

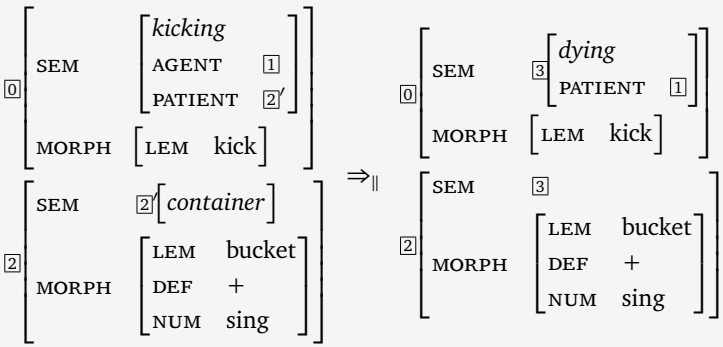

Figure 9 Global rewriting rule for *kick the bucket* that formalizes the quasi-inference rule in (17)

How can this be formalized? Actually, it is quite straightforward when reusing components of our lexical-semantic approach: the first step is to replace $\approx$ by $\Rightarrow_{\|}$, which is basically a rewriting rule that is defined as follows:

(18) $\quad a \Rightarrow_{\|} b \quad = \quad a \Rightarrow (a \,\|\, b)$

In prose, $\Rightarrow_{\|}$ takes the left-hand side (i.e., a) and wraps a $\|$-term around it with the right-hand side (i.e., b) being another $\|$-disjunct. Note that this is a very restricted, monotonic notion of term rewriting. The second step is to instantiate a and b in (18) with SEM-MORPH descriptions.

Treated in this way, the quasi-inference rule in (17) becomes the rewriting rule in figure 9, where the SEM-MORPH descriptions are directly taken from the lexical-semantic analysis in figure 3. Note that the role of Pulman's "lexical indexing," which serves to tie the quasi-inference rule to a surface string, is taken over by MORPH descriptions that are lumped together with components of SEM.

One virtue of this approach is immediately apparent: the idiomatic meaning can be condensed into one global rule, instead of flooding the lexicon with $\|$-terms. In other words, one can conceive the rewriting rule in figure 9 as a generalization over fully specified lexical entries of the lexical-ambiguity approach. Admittedly, this does not look spectacular in the example at hand, but note that this sort of rules can also be used for expressing much more generic, morphologically less fixed general-

izations, for example, conceptual metaphors, and inheritance relations among those.

So far, so good. There is, however, at least one aspect of Pulman's quasi-inference approach that does not seem to fit so neatly into the picture: quasi-inference rules are applied post-syntactically and rely on the complete instantiation of the left-hand side (plus fulfillment of lexical indexing). This means that Pulman assumes a strict two-step approach: first the literal meaning is computed, and only then is the idiomatic meaning deduced. However, this is in principle incompatible with lexical-semantic approaches, where literal and idiomatic interpretations are released in parallel. Furthermore, as laid out in section 3.2 already, the two-step approach has been criticized elsewhere for contradicting psycholinguistic findings (e.g., from Cacciari & Tabossi 1988) that suggest that idioms are processed incrementally, that is, approximately word by word. This means that even partial triggers of MWEs suffice to activate the idiomatic interpretation.

For this reason, it has to be taken into consideration whether quasi-inference rules, or ||-rewriting rules, should already apply based on incomplete left-hand sides. However, the effect would be that the conjectured computational advantage of inference-based approaches would disappear, because the idiomatic interpretation would be released on a per-word basis as well – or worse: one would have to add an extra distinction in order to specify the part of the left-hand side that has to be minimally present. This is necessary because we certainly don't want to allow for unmotivated, random applications. Similarly, one has to somehow prohibit infinite regress, that is, the recursive application of an inference rule to its right-hand side.

On the other hand, one possible advantage of quasi-inference rules, at least when it comes to the treatment of conjunctive modification, could be that they cleanly separate literal and idiomatic components. However, as we argued in section 4.3, this alone would not suffice anyway because it does not explain the possible co-existence of literal and idiomatic interpretations of the same phrase. Thus, in general, no substantial gain in coverage can be attested compared to lexical-semantic approaches.

Taken together, it seems to be preferable to do syntax within a lexical-semantic approach, while expressing lexical generalizations by means of

||-rewriting rules. The latter would be part of the metagrammar, but not immediately take part in parsing. Hence, under this view, the inference-based approach supplements the lexical-ambiguity approach rather than constitutes an alternative.

6 Conclusion

The aim of this work was fourfold: (i) to promote awareness of the sort of ambiguity that can emerge when dealing with semantically idiomatic MWEs; (ii) to argue in favor of semantic ambiguity approaches on psycholinguistic and computational grounds; (iii) to sketch a lexical-semantic approach based on LTAG, which improves on the partial-function approach of Gazdar et al. (1985); (iv) to entertain the idea that the inference-based approach of Pulman (1993), under a certain formalization, should be seen as a tool to express generalizations about the morphosemantic properties of lexical entries. Sure enough, we have barely touched upon these topics, and have skipped many others. Thus, the list of objects for future work is long, the most urgent ones being the integration of partial uses and conjunctive modification, and the explication of the meaning and treatment of the ambiguity operator ||. But we hope that the underlying ideas are clear.

Some readers might still be bothered that we haven't sufficiently limited the scope of our work, remaining rather silent as to how to model morphosyntactic flexibility, or other sorts of MWEs that it is supposed to cover. We have deliberately taken a semantic stance and concentrated on non-decomposable MWEs, hoping that it's obvious that decomposable MWEs such as light-verb constructions can be dealt with as well. Sorts of non-semantic idiomaticity, as we said, were left aside and could be treated the usual, that is, phrasal, way.

As far as this work is concerned, an important motivation was to explore ways of "graceful integration" (Jackendoff 2011) of grammar models with psycholinguistic findings about the mental processing of MWEs. From our point of view, this and the question of how to account for the figurative flexibility in MWEs deserve more attention and formally more explicit models.

Acknowledgments We are grateful to Sascha Bargmann, Benjamin Burkhardt,

Jamie Findlay, Christopher Piñón, Manfred Sailer, Christian Wurm, and an anonymous reviewer for detailed comments and discussion. We also thank the audience and reviewers of CSSP 2015 where a first version of this paper was presented. The initiative to work on this topic arose from the participation in the scientific network PARSEME (COST Action IC1207). The work presented in this paper was financed by the Deutsche Forschungsgemeinschaft (DFG) within the CRC 991 "The Structure of Representations in Language, Cognition, and Science."

References

Abeillé, Anne. 1990. Lexical and syntactic rules in a Tree Adjoining Grammar. In Robert C. Berwick (ed.), *Proceedings of the 28th Annual Meeting of the Association for Computational Linguistics (ACL '90)*, 292–298. Pittburgh, PA.

Abeillé, Anne. 1995. The flexibility of French idioms: A representation with Lexicalized Tree Adjoining Grammar. In Martin Everaert, Erik-Jan van der Linden, André Schenk & Rob Schreuder (eds.), *Idioms: Structural and psychological perspectives*, 15–42. Hillsdale, NJ: Lawrence Erlbaum.

Abeillé, Anne & Owen Rambow. 2000. Tree Adjoining Grammar: An overview. In Anne Abeillé & Owen Rambow (eds.), *Tree Adjoining Grammars: Formalisms, linguistic analyses and processing* (CSLI Lecture Notes 107), 1–68. Stanford, CA: CSLI Publications.

Abeillé, Anne & Yves Schabes. 1989. Parsing idioms in lexicalized TAGs. In Harold Somers & Mary McGee Wood (eds.), *Proceedings of the Fourth Conference of the European Chapter of the Association for Computational Linguistics (EACL '89)*, 1–9. Manchester, England.

Abeillé, Anne & Yves Schabes. 1996. Non-compositional discontinuous constituents in Tree Adjoining Grammar. In Harry Bunt & Arthur van Horck (eds.), *Discontinuous constituency*, 279–306. Mouton de Gruyter.

Baldwin, Timothy & Su Nam Kim. 2010. Multiword expressions. In Nitin Indurkhya & Fred J. Damera (eds.), *Handbook of natural language processing*, CRC Press 2nd edn.

Bresnan, Joan. 1982. The passive in lexical theory. In Joan Bresnan (ed.), *The mental representation of grammatical relations*, 40–65. Cambridge, MA: The MIT Press.

Cacciari, Cristina & Patrizia Tabossi. 1988. The comprehension of idioms. *Journal of Memory and Language* 27(6). 668–683.

Chomsky, Noam. 1980. *Rules and representations*. Basil Blackwell.

Crabbé, Benoît, Denys Duchier, Claire Gardent, Joseph Le Roux & Yannick Parmentier. 2013. XMG: eXtensible MetaGrammar. *Computational Linguistics*

39(3). 591–629.
Egan, Andy. 2008. Pretense for the complete idiom. *Noûs* 42(3). 381–409.
Ernst, Thomas. 1981. Grist for the linguistic mill: Idioms and 'extra' adjectives. *Journal of Linguistic Research* 1. 51–68.
Fischer, Ingrid & Martina Keil. 1996. Parsing decomposable idioms. In Jun-ichi Tsuji (ed.), *Proceedings of the 16th Conference on Computational Linguistics (COLING96)*, 388–393. Copenhagen, Denmark.
Gazdar, Gerald, Ewan Klein, Ivan A. Sag & Geoffrey K. Pullum. 1985. *Generalized Phrase Structure Grammar*. Harvard University Press.
Gibbs, Jr., Raymond W. 2002. A new look at literal meaning in understanding what is said and implicated. *Journal of Pragmatics* 34(4). 457–486.
Jackendoff, Ray. 2011. What is the human language faculty? Two views. *Language* 87(3). 586–624.
Joshi, Aravind K. & Yves Schabes. 1997. Tree-Adjoining Grammars. In G. Rozenberg & A. Salomaa (eds.), *Handbook of formal languages*, vol. 3, 69–124. Springer.
Kallmeyer, Laura & Rainer Osswald. 2013. Syntax-driven semantic frame composition in Lexicalized Tree Adjoining Grammar. *Journal of Language Modelling* 1. 267–330.
Kay, Paul, Ivan A. Sag & Dan Flickinger. in progress. A lexical theory of phrasal idioms. `http://www1.icsi.berkeley.edu/~kay/`.
Peterson, Robert R. & Curt Burgess. 1993. Syntactic and semantic processing during idiom comprehension: Neurolinguistic and psycholinguistic dissociations. In Christina Cacciari & Patrizia Tabossi (eds.), *Idioms: Processing, structure, and interpretation*, 201–225. Hillsdale, NJ: Lawrence Erlbaum.
Poesio, Massimo. 1996. Semantic ambiguity and perceived ambiguity. In Kees van Deemter & Stanley Peters (eds.), *Semantic ambiguity and underspecification*, 159–201. Stanford, CA: CSLI Publications.
Pulman, Stephen G. 1993. The recognition and interpretation of idioms. In Cristina Cacciari & Patrizia Tabossi (eds.), *Idioms: Processing, structure, and interpretation*, 249–270. Hillsdale, NJ: Lawrence Erlbaum.
Richter, Frank & Manfred Sailer. 2009. Phraseological clauses in constructional HPSG. In Stefan Müller (ed.), *Proceedings of the 16th International Conference on Head-driven Phrase Structure Grammar, University of Göttingen, Germany*, 297–317. Stanford, CA: CSLI Publications.
Riehemann, Susanne Z. 2001. *A constructional approach to idioms and word formation*. Stanford, CA: Stanford University Dissertation. `http://doors.stanford.edu/~sr/sr-diss.pdf`.

Récanati, François. 1995. The alleged priority of literal interpretation. *Cognitive Science* 19(2). 207–232.

Sag, Ivan A., Timothy Baldwin, Francis Bond, Ann Copestake & Dan Flickinger. 2002. Multiword expressions: A pain in the neck for NLP. In Alexander Gelbukh (ed.), *Computational linguistics and intelligent text processing* (Lecture Notes in Computer Science 2276), 1–15. Springer.

Sailer, Manfred. 2003. *Combinatorial semantics and idiomatic expressions in Head-Driven Phrase Structure Grammar* (Arbeitspapiere des SFB 340 161). Tübingen, Germany: Universität Tübingen. http://hdl.handle.net/10900/46191.

Soehn, Jan-Philipp. 2006. *Über Bärendienste und erstaunte Bauklötze: Idiome ohne freie Lesart in der HPSG*. Frankfurt a. M.: Peter Lang.

Wearing, Catherine. 2012. Metaphor, idiom, and pretense. *Noûs* 46(3). 499–524.

Wittenberg, Eva, Ray Jackendoff, Gina Kuperberg, Martin Paczynski, Jesse Snedeker & Heike Wiese. 2014. The processing and representation of light verb constructions. In Asaf Bachrach, Isabelle Roy & Linnaea Stockall (eds.), *Structuring the argument*, 61–80. John Benjamins.

Wittenberg, Eva & Maria Mercedes Piñango. 2011. Processing light verb constructions. *The Mental Lexicon* 6(3). 393–413.

Wurm, Christian & Timm Lichte. 2016. The proper treatment of linguistic ambiguity in ordinary algebra. In Annie Foret, Glyn Morrill, Reinhard Muskens, Rainer Osswald & Sylvain Pogodalla (eds.), *Formal Grammar: 20th and 21st International Conferences, FG 2015, Barcelona, Spain, August 2015, Revised selected papers. FG 2016, Bozen, Italy, August 2016, Proceedings* (Lecture Notes in Computer Science 9804), 306–322. Berlin: Springer.