(WG 1 and WG 2)

Two ways of modelling idiomaticity as semantic ambiguity in LTAG

Timm Lichte
CRC 991, University of Düsseldorf
lichte@phil.hhu.de

Laura Kallmeyer
CRC 991, University of Düsseldorf
kallmeyer@phil.hhu.de

Abstract Idiomatic MWEs are commonly analysed 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 two alternatives, employing the framework of LTAG, where idiomaticity is not subject to syntactic ambiguity, but emerges in the semantics.

Idiomaticity as syntactic idiomaticity Lexicalized Tree-Adjoining Grammar (LTAG, [4]) is renowned to provide elegant accounts to a range of multi-word expressions with non-compositional meaning (e. g. [1]). 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). An example is shown in Figure 1 with a frame-based semantics following [5]. Due to the flexible linking of syntax and semantics by means of interface variables (see boxed numbers), internal and external modification can be adequately handled.

However, the framework tends to enumerate non-compositional expressions by assigning a separate elementary tree to each of them. While this syntactic ambiguity can be dealt with on the technical side by means of metagrammar tools (e. g. XMG), further desiderata remain: the approach rather neglects recent psycholinguistic findings about the processing of idiomatic MWEs [7], and it obscures the possible connection between their literal and idiomatic meaning.

Idiomaticity as semantic ambiguity Taking up work in [3], we are currently exploring ways to implement compositional idiomaticity as a result of semantically determined ambiguity without the sharing of syntax. Two rather distinct options seem to be available: semantic ambiguity induced by lexical specification or by global entailments.

The first option, lexically induced ambiguity, is sketched in Figure 2. Here each elementary tree comes with a disjunction of morphologically enriched frame specifications. While special elementary trees for MWEs are missing, the morphological features are necessary for confining the context where the idiomatic meaning emerges. Note two other aspects of this approach: the frame of the verbal head must be visible in the NP-slot; and the readings of a lexeme are immediately available when instantiating it's elementary tree. While the first aspect seems unproblematic, the second could be taken to contravene psycholinguistic results that suggest that readings are not equally available. Putting distinct weights on the disjuncts might solve this issue.

The second option based on global entailments is exemplified in Figure 3. These entailments are global in the sense that they in principle apply independently from the lexicon or the syntactic derivation: the structure matching the left hand side is minimally changed or "forced" to entail also the right hand side. Hence they are non-monotonic and bear formal similarity to lexical rules or unary grammar rules in HPSG [2, 6]. Despite the well-known computational drawbacks of non-monotonicity, this approach also has its virtues: the idiomatic meaning can be condensed in one global entailment and the emergence of readings can be delayed. Still it is unclear, whether non-monotonicity is really necessary.

Besides the elaboration of their formal details, we are interested in how the two approaches cope with the pronominalization and isolation of MWE components, and how they carry over to other sorts of MWEs. Only after clarifying these questions can the two approaches be reasonably compared.

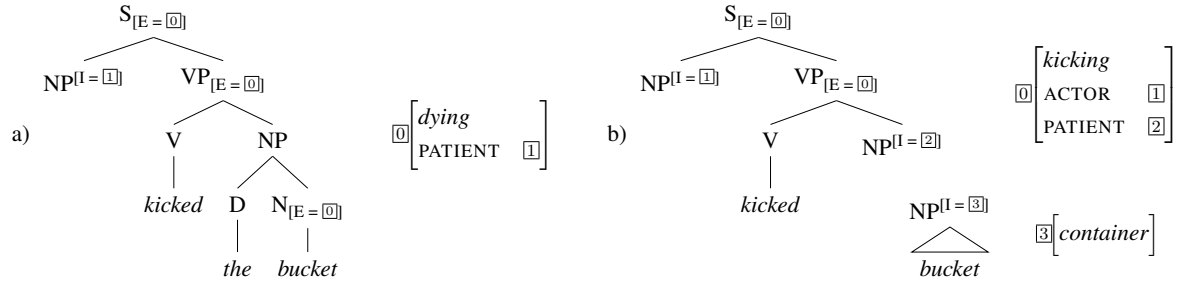

Figure 1: Syntactic ambiguity induced by disjunction over pairs of elementary trees and frame semantic representations.

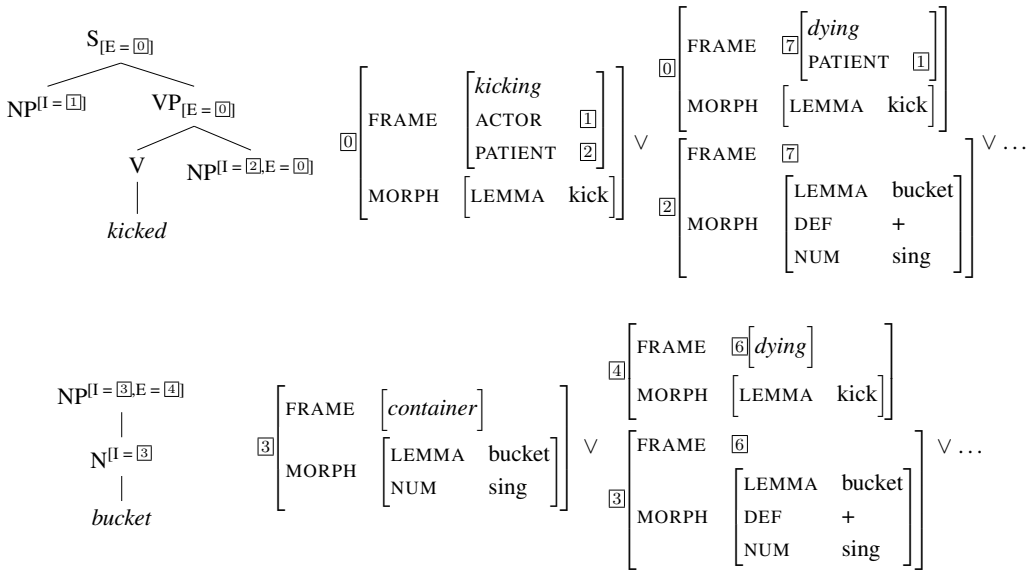

Figure 2: Semantic ambiguity due to lexically induced disjunction over frame semantic and morphological structures.

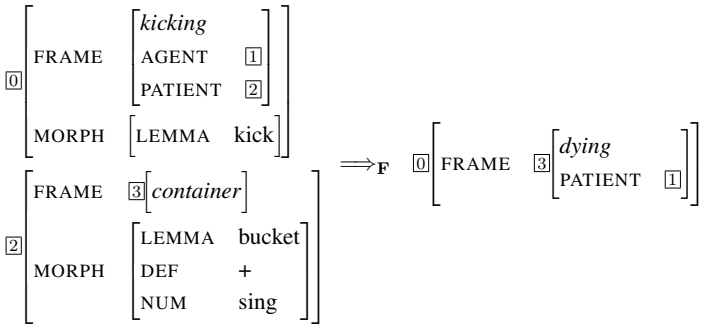

Figure 3: Semantic ambiguity due to global, non-monotonic force entailments over frame semantic and morphological structures.

References

[1] Abeillé, Anne & Yves Schabes. 1996. Non-compositional discontinuous constituents in tree adjoining grammar. In Harry Bunt & Arthur van Horck (eds.), *Discontinuous constituency*, 279–306. Berlin, New York: Mouton de Gruyter.

[2] Briscoe, Ted & Ann Copestake. 1999. Lexical rules in constraint-based grammars. *Computational Linguistics* 25(4). 487–526.

[3] 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.

[4] Joshi, Aravind K. & Yves Schabes. 1997. Tree-Adjoining Grammars. In G. Rozenberg & A. Salomaa (eds.), *Handbook of formal languages*, vol. 3, 69–124. Berlin, New York: Springer.

[5] Kallmeyer, Laura & Rainer Osswald. 2013. Syntax-driven semantic frame composition in Lexicalized Tree Adjoining Grammar. *Journal of Language Modelling* 1. 267–330.

[6] Meurers, W. Detmar. 2000. *Lexical generalizations in the syntax of German non-finite constructions*. Tübingen, Germany: Seminar für Sprachwissenschaft, Universität Tübingen dissertation. Published 2000 as Volume 145 in Arbeitspapiere des SFB 340, ISSN 0947-6954/00.

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