

# Levels of ambiguity and underspecification in regular polysemy

Laura Kallmeyer<sup>1</sup>, Rainer Osswald<sup>1</sup> & Christian Retoré<sup>2</sup>

<sup>1</sup>Heinrich-Heine-Universität Düsseldorf

<sup>2</sup>Université de Montpellier & LIRMM

It is common to contrast ambiguity with context sensitivity and underspecification (Sennet, 2016): While the context may help to disambiguate an ambiguous expression, the ambiguity of the expression does not depend on the context. That is, the different possible interpretations of an ambiguous expression are given independently of the contexts in which it occurs. Underspecification, on the other hand, differs from ambiguity in that a lack of specificity does not imply that the expression in question has different, clearly distinguishable interpretations. Polysemy, by comparison, is generally considered as falling between ambiguity and vagueness (Tuggy, 1993): As with ambiguity, the different meanings are clearly distinguishable, as with vagueness, they are clearly related to each other. What does this mean for the relation of polysemy to context sensitivity and underspecification? On the one hand, polysemy is often treated as a special case of ambiguity (Sennet, 2016), thereby implying a contrast to context sensitivity and underspecification. On the other hand, contextual information and underspecification play a prominent role in various approaches to the modeling of regular (or logical or systematic) polysemy (e.g. Egg 2003, 2011; Asher 2011; Mery and Retoré 2015; Babonnaud et al. 2016). The goal of the present paper is to explore the role of ambiguity versus underspecification and context dependency in the analysis of polysemy.

Pustejovsky (2011) draws a distinction between inherent and selectional regular polysemy. Inherent polysemy means that different facets of interpretation are inherent in the lexical semantics of a word. Standard examples are *book* and *lunch*, which can denote physical or informational objects, and food or events, respectively. Selectional polysemy refers to cases of apparent selectional mismatches between predicate and argument which are resolved by coercion mechanisms that go beyond inherent meaning shifts. Examples are *Mary left the party* and *Mary began the book*, where *leave* selects for a location argument and *begin* selects for an event argument. In which sense can we speak of polysemy in these cases? Asher (2011), among others, holds that neither the meaning of the argument nor that of the predicate is shifted or changed; what is modified is rather the predication relation (the type presupposition on the argument). Hence, none of the lexical items involved is required to be polysemous and neither is their combination. However, the combined expressions are often ambiguous due to possible variations in the coercion process. While in the case of *leave the party*, the shift from event to location does not allow much room for ambiguity, the expression *begin the book* is ambiguous with respect to which kind of interaction with the book (reading, writing, translating, memorizing, coloring, mending, etc.) is being started. Note that this is not an instance of underspecification since the type of interaction meant by the expression is usually clear in a given context. In particular, an explicitly underspecified formulation such as *Mary began to do something related to the book* is not a legitimate paraphrase of *Mary began the book*. If underspecification plays a role in the analysis of selectional polysemy then necessarily at some intermediate level of the coercion process. Moreover, it seems problematic to classify examples like *begin the book* as polysemous since the different activities involving the book do neither form a circumscribed set of clearly distinguishable alternatives (even if some of them are more salient than others) nor are they clearly related to each other, except for the fact that they potentially involve books.

For a more thorough discussion of these issues, we make use of a formal model of the syntax-semantics interface which combines Lexicalized Tree Adjoining Grammars (LTAG) with frame semantics (Kallmeyer and Osswald, 2013; Babonnaud et al., 2016). In this framework, the basic components are elementary constructions: syntactic trees paired with semantic frames and frame descriptions, in which interface features relate syntactic components to com-

ponents of the semantic structures. Semantic composition is driven by tree substitution and adjunction.

Any formal analysis of inherent polysemy has to cope with two questions: how to represent the different meaning facets, and how to access them selectively depending on the context. The lexical-semantic representations proposed in the literature can be distinguished according to whether or not they postulate a complex type of entity for simultaneously encoding the different meaning facets. In either case, polysemous lexical entries are assumed to contain all of their facets (in contrast to homographs or homophones, which are stored in separate entries). Since the different meaning facets of a polysemous expression are related, the underlying relations should be part of the semantic representation. The simple example in (1) illustrates how this can be implemented in the LTAG + frame semantics approach without invoking complex, multi-facet entities.

$$(1) \quad \begin{array}{ccc} \text{phys-obj} & \xrightarrow{\text{information}} & \text{information} \\ \textcircled{u} & \xrightarrow{\text{CONTENT}} & \textcircled{\phantom{u}} \\ & & \text{NP}^{[I=x]} \\ & & | \\ & & \text{'book'} \end{array} \quad x \triangleq u \vee x \triangleq u \cdot \text{CONTENT}$$

The analysis of books as physical information carriers is represented by the frame on the left of (1), which is an instance of the frame type associated with the description  $\text{phys-obj} \wedge \text{CONTENT}:\text{information}$ . The relation between the meaning facets is represented by the frame attribute CONTENT while the referential index  $u$  at the syntax-semantics interface can target the different alternatives. The disjunction is exclusive because of the type constraints associated with CONTENT. Disambiguation can then be enforced by unification with appropriate selectional restrictions provided by a verb or modifier. A problem of the representation in (1) is that the ambiguity is bound to instances, whereas, in fact, a type-level ambiguity is required in general. For example, a quantified expression such as *every book on the shelf* should refer to all physical books or to all book contents. One way to cope with this issue is to encode the ambiguity via underspecification at the level of frame descriptions (along the lines of Babonnaud et al. 2016; Kallmeyer and Osswald 2017). The basic idea is sketched in (2):

$$(2) \quad \begin{array}{ccc} l_1 : \exists u(u \cdot \textcircled{1}), & \textcircled{1} \triangleleft^* l_2, & \text{NP}^{[P=\textcircled{1}]} \\ l_2 : \text{phys-obj} \wedge \text{CONTENT}:\textcircled{2}, & \textcircled{2} \triangleleft^* l_3, & | \\ l_3 : \text{information} & & \text{NP}^{[I=\textcircled{0}], P=l_2, \text{TOP}=l_1} \\ & & | \\ & & \text{'book'} \end{array}$$

Here, the frame description is split into components in order to allow, in combination with the specification of the quantifier, the quantified variable to occur at different but fixed positions in the formula. ( $h \triangleleft^* l$  means that expression  $l$  is a subexpression of  $h$ .) Details will be given in the talk. Note that the underspecification approach just described is fully compatible with the view on ambiguity introduced above since it basically consists of a system of constraints on the form of the logical structure which comes with the requirement of resolving these constraints in one way or another.

Selectional polysemy, as explained above, is a misnomer for argument coercion, which can, but need not give rise to ambiguity. Consider again the expression *leave the party*. The sketch of the entry of *leave* in (3) indicates that the location  $u$  that is left is either the entity  $y$  denoted by the object NP or the LOCATION of  $y$ .

$$(3) \quad \begin{array}{c} e \cdot (\text{leaving} \wedge \text{AGENT}:x \wedge \text{THEME}:(u \cdot \text{location})) \\ \wedge (u \triangleq y \vee u \triangleq y \cdot \text{LOCATION}) \end{array} \quad \begin{array}{c} \text{S}^{[I=e]} \\ \hline \text{NP}^{[I=x]} \quad \text{VP} \\ \hline \text{V} \quad \text{NP}^{[I=y]} \\ | \\ \text{'leave'} \end{array}$$

The disjunction in (3) should be regarded as a simple distinction of cases and not as a sign of ambiguity or underspecification. In the talk, we will show to what extent this is different for the *begin the book* type of examples.

## References

- Asher, N. (2011). *Lexical Meaning in Context. A Web of Words*. Cambridge University Press, Cambridge.
- Babonnaud, W., Kallmeyer, L., and Osswald, R. (2016). Polysemy and coercion – a frame-based approach using LTAG and Hybrid Logic. In Amblard, M., de Groote, P., Pogodalla, S., and Retoré, C., editors, *Logical Aspects of Computational Linguistics, 9th International Conference*, Lecture Notes in Artificial Intelligence 10054, pages 18–33. Springer, Berlin.
- Egg, M. (2003). Beginning novels and finishing hamburgers: Remarks on the semantics of ‘to begin’. *Journal of Semantics*, 20:163–191.
- Egg, M. (2011). Semantic underspecification. In Maienborn, C., von Heusinger, K., and Portner, P., editors, *Semantics. An International Handbook of Natural Language Meaning. Volume 1*, pages 535–574. Mouton de Gruyter, Berlin.
- Kallmeyer, L. and Osswald, R. (2013). Syntax-driven semantic frame composition in Lexicalized Tree Adjoining Grammars. *Journal of Language Modelling*, 1(2):267–330.
- Kallmeyer, L. and Osswald, R. (2017). Modeling quantification with polysemous nouns. In Gardent, C. and Retoré, C., editors, *IWCS 2017 – 12th International Conference on Computational Semantics – Short papers*. ACL Anthology.
- Mery, B. and Retoré, C. (2015). Are books events? Ontological inclusions as coercive sub-typing, lexical transfers as entailment. In *Proceedings of the 12 Workshop on Logic and Engineering of Natural Language Semantics 12 (LENLS12)*, pages 74–87, Tokyo.
- Pustejovsky, J. (2011). Coercion in a general theory of argument selection. *Linguistics*, 49(6):1401–1431.
- Sennet, A. (2016). Ambiguity. In Zalta, E. N., editor, *The Stanford Encyclopedia of Philosophy*. Metaphysics Research Lab, Stanford University, spring 2016 edition.
- Tuggy, D. (1993). Ambiguity, polysemy, and vagueness. *Cognitive Linguistics*, 4(3):273–290.