

Interlingual Meaning Representations

Johan Bos

University of Groningen
johan.bos@rug.nl

What is *meaning*, and how can it be captured in a concrete representation? This is a challenging question, given the fact that meanings enjoy a large degree of abstraction. A question that directly follows from this is to what extent meanings (or if you prefer: *representations of meaning*), need to be language neutral. Most of the current large-scale meaning representations employed in natural language processing are tailored to specific object languages, often English [4, 5, 3] because it has been the dominating language of study in computational linguistics (but this is slowly changing, fortunately). This is completely understandable from a short-term, practical perspective. But from a theoretical point of view, this doesn't make sense at all (just think about how translations from one language into another preserve meaning). A natural question to ask, then, is how far we can stretch *interlingual* meaning representations. What is required to achieve this—what resources and (linguistic) knowledge do we need? What challenges are we facing? What role can and must logic play?

In most logical approaches to semantics, a part of the meaning representation is, by its very nature, independent of the object language: the logical symbols used to express negation, conjunction, disjunction, and quantification. The non-logical symbols are usually represented by strings resembling words of a specific language (again, usually English). This is a tradition started by Montague [11], and followed by many others [7, 8]. So logic only gives a partial guidance to our endeavour of making meaning representations more interlingual. Should we expect more from logic? What is a good balance between logical and non-logical ingredients in a meaning representation? Let us look at a concrete example.

The Parallel Meaning Bank, PMB [1], is a semantically annotated corpus for four languages (English, Dutch, German, and Italian). It comprises translations between these languages, and under the assumption that translations preserve meaning, the PMB is the perfect environment to investigate interlingual meaning representations. The meaning representations in the PMB combine the logical aspects of Discourse Representation Theory [9] with lexical resources including WordNet [6], VerbNet [10], and FrameNet [2]. The PMB data demonstrates that even closely related languages behave differently in (for instance) marking definiteness, realisation of verbal arguments, and multi-word expressions. Despite these new challenges, I will argue that providing interlingual meaning representations is a welcome direction not only in computational, but also in formal approaches to meaning.

References

1. Abzianidze, L., Bjerva, J., Evang, K., Haagsma, H., van Noord, R., Ludmann, P., Nguyen, D.D., Bos, J.: The parallel meaning bank: Towards a multilingual corpus of translations annotated with compositional meaning representations. In: Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics. pp. 242–247. Valencia, Spain (2017)
2. Baker, C.F., Fillmore, C.J., Lowe, J.B.: The Berkeley FrameNet project. In: 36th Annual Meeting of the Association for Computational Linguistics and 17th International Conference on Computational Linguistics. Proceedings of the Conference. pp. 86–90. Université de Montréal, Montreal, Quebec, Canada (1998)
3. Banarescu, L., Bonial, C., Cai, S., Georgescu, M., Griffitt, K., Hermjakob, U., Knight, K., Koehn, P., Palmer, M., Schneider, N.: Abstract Meaning Representation for Sembanking. In: Proceedings of the 7th Linguistic Annotation Workshop and Interoperability with Discourse. pp. 178–186. Sofia, Bulgaria (August 2013), <http://www.aclweb.org/anthology/W13-2322>
4. Bos, J., Basile, V., Evang, K., Venhuizen, N., Bjerva, J.: The Groningen Meaning Bank. In: Ide, N., Pustejovsky, J. (eds.) Handbook of Linguistic Annotation, vol. 2, pp. 463–496. Springer (2017)
5. Butler, A.: The Semantics of Grammatical Dependencies, vol. 23. Emerald Group Publishing Limited (2010)
6. Fellbaum, C. (ed.): WordNet. An Electronic Lexical Database. The MIT Press (1998)
7. Heim, I.: The Semantics of Definite and Indefinite Noun Phrases. Ph.D. thesis, University of Massachusetts (1982)
8. Kamp, H.: A Theory of Truth and Semantic Representation. In: Groenendijk, J., Janssen, T.M., Stokhof, M. (eds.) Truth, Interpretation and Information, pp. 1–41. FORIS, Dordrecht – Holland/Cinnaminson – U.S.A. (1984)
9. Kamp, H., Reyle, U.: From Discourse to Logic; An Introduction to Modeltheoretic Semantics of Natural Language, Formal Logic and DRT. Kluwer, Dordrecht (1993)
10. Kipper, K., Korhonen, A., Ryant, N., Palmer, M.: A large-scale classification of English verbs. *Language Resources and Evaluation* **42**(1), 21–40 (2008)
11. Montague, R.: The proper treatment of quantification in ordinary English. In: Hintikka, J., Moravcsik, J., Suppes, P. (eds.) Approaches to Natural Language, pp. 221–242. Reidel, Dordrecht (1973)