

Towards a Better Semantic Role Labeling of Complex Predicates

Poster Proposal concerning WG 4 or 3

1 Introduction

The work presented on this poster focuses on complex predicates (CPs). More in particular, we propose a method that enables a better handling of CPs for the task of Semantic Role Labeling (SRL) based on the lexical resource PropBank (PB) (Palmer et al., 2005). As Bonial et al. (2014) recently stated ‘PB has previously treated language as if it were purely compositional, and has therefore lumped the majority of MWEs in with lexical verb usages’. For example the predicates in the CPs *take a hard line*, *take time* and many others are annotated with a sense of *take*, meaning *acquire*, *come to have*, *chose*, *bring with you from somewhere*. This results in a loss of semantic information in PB annotations and automatic SRL systems trained on this data.

Recently, Bonial et al. (2014) have introduced an approach to better handle MWEs in PB. The process is called *aliasing*. Instead of creating new frames for CPs, human annotators map them to existing PB rolesets which encompass the same semantics and argument structure. For example, the alias *lecture.01* is mapped to the CP *give (a) talk*. Our work extends on this approach by finding aliases automatically thereby alleviating the need for human annotation. We use a multilingual distributional method to find appropriate aliases for two types of English CPs: light verb construction (LVC) and verb particle construction (VPC).

2 Methods

The task of finding aliases for CPs automatically is related to finding (near-)synonymous rolesets for the CPs. We apply the distributional hypothesis (Harris, 1968) to the multilingual context. The intuition behind this is that verbs and CPs that are translated with similar words in many language are likely to be near-synonyms and therefore good candidate aliases. We use the translations of the CPs and verbs labeled with PB rolesets to 20 languages, as provided by the word alignments in the multilingual parallel corpus Europarl (Koehn, 2005). Note that, contrary to traditional work on multilingual distributional semantics, we exploited the rolesets assigned to verbs. Therefore, the near-synonyms identified by our system are specific to a certain sense of a verb and not only to its lexical form.

3 Evaluation framework

In order to evaluate our system, we set up an annotation effort. We selected 50 LVCs and 50 VPCs from the Wiki50 corpus (Vincze et al., 2011) and made sure that the test set was balanced for frequency. Four annotators were presented with the CP in context and were asked to propose one or several PB aliases which encompass the same meaning and argument structure. One annotator (A) labeled the whole set of 100 expressions; the three other annotators (B,C,D) each labeled one third of the expressions assigned randomly, so that every expression was annotated by two annotators. All occurrences of these 100 CP types were annotated, which resulted in a total of 197 annotated instances. The simple inter-annotator agreement was 67% for annotator A&B, 51% for A&C and 44% for A&D. The annotations were discussed among the annotators in order to reach a final agreed-upon test set. CPs which were judged semantically compositional or for which no

alias could be found were discarded. The final Wiki50 set consists of 154 instances of CPs labeled with one to four appropriate PB alias rolesets.

We compared our system with an informed baseline system that distinguishes between VPCs and LVCs. For VPCs, it checks whether there exists a PB multiword predicate for the expression and selects the first roleset of this predicate (e.g. there exists a predicate called *open_up* (*open.03*) for the VPC ‘open up’). For LVCs, it checks whether the noun has a corresponding verb predicate in PB and selects the first roleset of this predicate (e.g. *walk.01* for *take a walk*). Note that this informed baseline is very hard to beat and only fails in case of lack in coverage.

4 Results and Discussion

We evaluated our approach on the 154 CPs annotated in the course of this work (Wiki50 set), as well as on 70 CPs stemming from the annotations of Bonial et al. (2014). Our system outperforms the baseline in terms of coverage by a large margin. In terms of accuracy it outperforms the baseline for the subset of expressions for which no multiword PB predicate exists (36.6% vs. 34.0%). These numbers seem low, but note that this evaluation is very strict and rejects synonymous rolesets. We also applied a lenient evaluation that accepts aliases that are grouped under the same VerbNet class (Kipper-Schuler, 2006) as the alias provided in the gold standard, which results in an accuracy of 60% on the same subset. In general, the results are not that far from the upper bounds from human annotation (between 44% and 67%). These results suggest that the method could alleviate the manual annotation task of finding aliases for CPs that are not yet covered by PB.

The results on the take set from Bonial et al. (2014) were substantially lower, but, in contrast to the Wiki50 set, this test set is more restricted and less natural, because it contains only expressions that comprise the verb *take* and is partly derived from lexical resources instead of corpora.

5 Conclusions

We have presented an approach to handle CPs in SRL, that extends the work by Bonial et al. (2014) by automatically linking VPCs and LVCs to the PB roleset that best describes their meaning. Because we do not rely on manual annotations, we managed to scale up previous work, that only annotated few examples of *take* CPs. Instead, we rely on multilingual word-alignments in a parallel corpus automatically annotated on the English side and multilingual distributional methods. Results suggest that the method could alleviate the manual annotation task of finding aliases for CPs that are not yet covered by PB. In addition, we set up an annotation effort to gather a frequency-balanced, naturally occurring evaluation set that is more varied and larger than the annotations provided by Bonial et al. (2014).

References

- Bonial, C., Green, M., Preciado, J., and Palmer, M. (2014). An approach to take multi-word expressions. In *Proceedings of the 10th Workshop on Multiword Expressions (MWE)*, pages 94–98, Gothenburg, Sweden. Association for Computational Linguistics.
- Harris, Z. S. (1968). *Mathematical Structures of Language*. Wiley, New York, NY, USA.
- Kipper-Schuler, K. (2006). *VerbNet: A Broad-Coverage, Comprehensive Verb Lexicon*. PhD thesis, University of Pennsylvania.
- Koehn, P. (2005). Europarl: A Parallel Corpus for Statistical Machine Translation. In *Conference Proceedings: the tenth Machine Translation Summit*, pages 79–86, Phuket, Thailand. AAMT, AAMT.

- Palmer, M., Gildea, D., and Kingsbury, P. (2005). The proposition bank: An annotated corpus of semantic roles. *Comput. Linguist.*, 31(1):71–106.
- Vincze, V., Nagy T., I., and Berend, G. (2011). Multiword expressions and named entities in the wiki50 corpus. In *Proceedings of the International Conference Recent Advances in Natural Language Processing 2011*, pages 289–295, Hissar, Bulgaria. RANLP 2011 Organising Committee.