Beyond Multiword Expressions: Processing Idioms and Metaphors

Valia Kordoni
Humboldt-Universität zu Berlin (Germany)
kordonie@anglistik.hu-berlin.de

1 Introduction

Idioms and metaphors are characteristic to all areas of human activity and to all types of discourse. Their processing is a rapidly growing area in NLP, since they have become a big challenge for NLP systems. Their omnipresence in language has been established in a number of corpus studies and the role they play in human reasoning has also been confirmed in psychological experiments. This makes idioms and metaphors an important research area for computational and cognitive linguistics, and their automatic identification and interpretation indispensable for any semantics-oriented NLP application.

This tutorial aims to provide attendees with a clear notion of the linguistic characteristics of *idioms* and *metaphors*, computational models of *idioms* and *metaphors* using state-of-the-art NLP techniques, their relevance for the intersection of deep learning and natural language processing, what methods and resources are available to support their use, and what more could be done in the future. Our target audience are researchers and practitioners in machine learning, parsing (syntactic and semantic) and language technology, not necessarily experts in *idioms* and *metaphors*, who are interested in tasks that involve or could benefit from considering *idioms* and *metaphors* as a pervasive phenomenon in human language and communication.

This tutorial consists of four parts. Part I starts with an introduction to MWEs and their linguistic dimensions, that is, idiomaticity, syntactic and semantic fixedness, specificity, etc., as well as their statistical characteristics (variability, recurrence, association, etc.). The second half of this part focuses on the specific characteristics of idioms and metaphors (linguistic, conceptual and extended metaphor).

Part II surveys systems for processing idioms and metaphors which incorporate state-of-the-art NLP methods. The second half of this part is dedicated to resources for idioms and metaphors, as well as evaluation.

Part III offers a thorough overview of how and where research on idioms and metaphors can contribute to the intersection of NLP and Deep Learning, particularly focusing on recent advances in the computational treatment of MWEs in the framework of Deep Learning.

Part IV of the tutorial concludes with concrete examples of where idioms and metaphors treatment can contribute to language technology applications such as sentiment analysis, educational applications, dialog systems and digital humanities.

2 Tutorial Outline

1. PART I – General overview:
   (a) Introduction to MWEs: linguistic dimensions (idiomaticity, syntactic and semantic fixedness, specificity, etc.) and statistical dimensions (variability, recurrence, association, etc.)
   (b) Linguistic characteristics of idioms
   (c) Linguistic characteristics of metaphors (linguistic, conceptual and extended metaphor)
2. PART II – Systems for processing idioms and metaphors, resources and evaluation
   (a) Machine learning for idioms and metaphors
   (b) Generation of idioms and metaphors
   (c) Multilingual processing and translation of idioms and metaphors
   (d) Annotation of idioms and metaphors in corpora
   (e) Idioms and metaphors in lexical resources

Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics-Tutorial Abstracts, pages 15–16
Melbourne, Australia, July 15 - 20, 2018.

(f) Evaluation methodologies and frameworks

3. PART III – At the intersection of Deep learning and NLP
   (a) Beyond learning word vectors
   (b) Recursive Neural Networks for parsing idioms and metaphors
4. PART IV – Resources and applications:
   (a) Idioms and metaphors in Language Technology applications: sentiment analysis, educational applications, dialog systems and digital humanities

3 Tutorial Instructor

Valia Kordoni is a professor at Humboldt University Berlin (Deputy Chair for the subject area "English Linguistics"). She is a leader in EU-funded research in Machine Translation, Computational Semantics, and Machine Learning. She has organized conferences and workshops dedicated to research on MWEs, recently including the EACL 2014 *10th Workshop on Multiword Expressions (MWE 2014)* in Gothenburg, Sweden, the NAACL 2015 *11th Workshop on Multiword Expressions* in Denver, Colorado, and the ACL 2016 *12th Workshop on Multiword Expressions* in Berlin, Germany, among others. She has been the Local Chair of *ACL 2016 - The 54th Annual Meeting of the Association for Computational Linguistics* which took place at the Humboldt University Berlin in August 2016. Recent activities of hers include a tutorial on *Robust Automated Natural Language Processing with Multiword Expressions and Collocations* in ACL 2013, as well as a tutorial on *Beyond Words: Deep Learning for Multiword Expressions and Collocations* in ACL 2017. She is the author of *Multiword Expressions - From Linguistic Analysis to Language Technology Applications* (to appear, Springer).

References

Samuel R. Bowman, Christopher Potts, and Christopher D. Manning. 2015a. Learning distributed word representations for natural logic reasoning. In *Proceedings of the AAAI Spring Symposium on Knowledge Representation and Reasoning*.

Samuel R. Bowman, Christopher Potts, and Christopher D. Manning. 2015b. Recursive neural networks can learn logical semantics. In *Proceedings of the 3rd Workshop on Continuous Vector Space Models and their Compositionality*.

Danqi Chen and Christopher Manning. 2014. A fast and accurate dependency parser using neural networks. In *EMNLP*.

Spence Green, Marie-Catherine de Marneffe, and Christopher D. Manning. 2013. Parsing models for identifying multiword expressions. *Computational Linguistics*, 39(1):195–227.

Eric H. Huang, Richard Socher, Christopher D. Manning, and Andrew Y. Ng. 2012. Improving word representations via global context and multiple word prototypes. In *ACL*.

Su Nam Kim and Preslav Nakov. 2011. Large-scale noun compound interpretation using bootstrapping and the web as a corpus. In *EMNLP*, pages 648–658.

Beata Beigman Klebanov, Ekaterina Shutova, and Patricia Lichtenstein, editors. 2014. *Proceedings of the Second Workshop on Metaphor in NLP*. Association for Computational Linguistics, Baltimore, MD, June.

Christopher D. Manning and Hinrich Schütze. 2001. *Foundations of statistical natural language processing*. MIT Press.

Carlos Ramisch, Aline Villavicencio, and Valia Kordoni. to appear. *Special Issue on Multiword Expressions*. ACM TSLP.

Paul Rayson, Scott Songlin Piao, Serge Sharoff, Stefan Evert, and Begoña Villada Moirón. 2010. Multiword expressions: hard going or plain sailing? *Language Resources and Evaluation*, 44(1-2):1–5.

Ivan A. Sag, Timothy Baldwin, Francis Bond, Ann Copestake, and Dan Flickinger. 2001. Multiword expressions: A pain in the neck for nlp. In *In Proc. of the 3rd International Conference on Intelligent Text Processing and Computational Linguistics (CICLing-2002*, pages 1–15.

Violeta Seretan. 2012. *Syntax-Based Collocation Extraction*, volume 44, Text, Speech and Language Technology. Springer.

Ekaterina Shutova, Beata Beigman Klebanov, Joel Tetreault, and Zornitsa Kozareva, editors. 2013a. *Proceedings of the First Workshop on Metaphor in NLP*. Association for Computational Linguistics, Atlanta, Georgia, June.

Ekaterina Shutova, Simone Teufel, and Anna Korhonen. 2013b. Statistical metaphor processing. *Comput. Linguist.*, 39(2):301–353, June.

Aline Villavicencio, Francis Bond, Anna Korhonen, and Diana McCarthy. 2005. Introduction to the special issue on multiword expressions: Having a crack at a hard nut. *Computer Speech & Language*, 19(4):365–377.