



Contribution to Semantic Parsing Approaches and Techniques



Gregorio Nuevo Castro
Supervised by: Felipe Gil Castiñeira¹

¹Department of Telematics Engineering, University of Vigo

Introduction

Syntactic Parsing – extracting syntactic structure from text – is an old task of **Natural Language Processing** (NLP) and has received great improvements over last decades (figure 1).

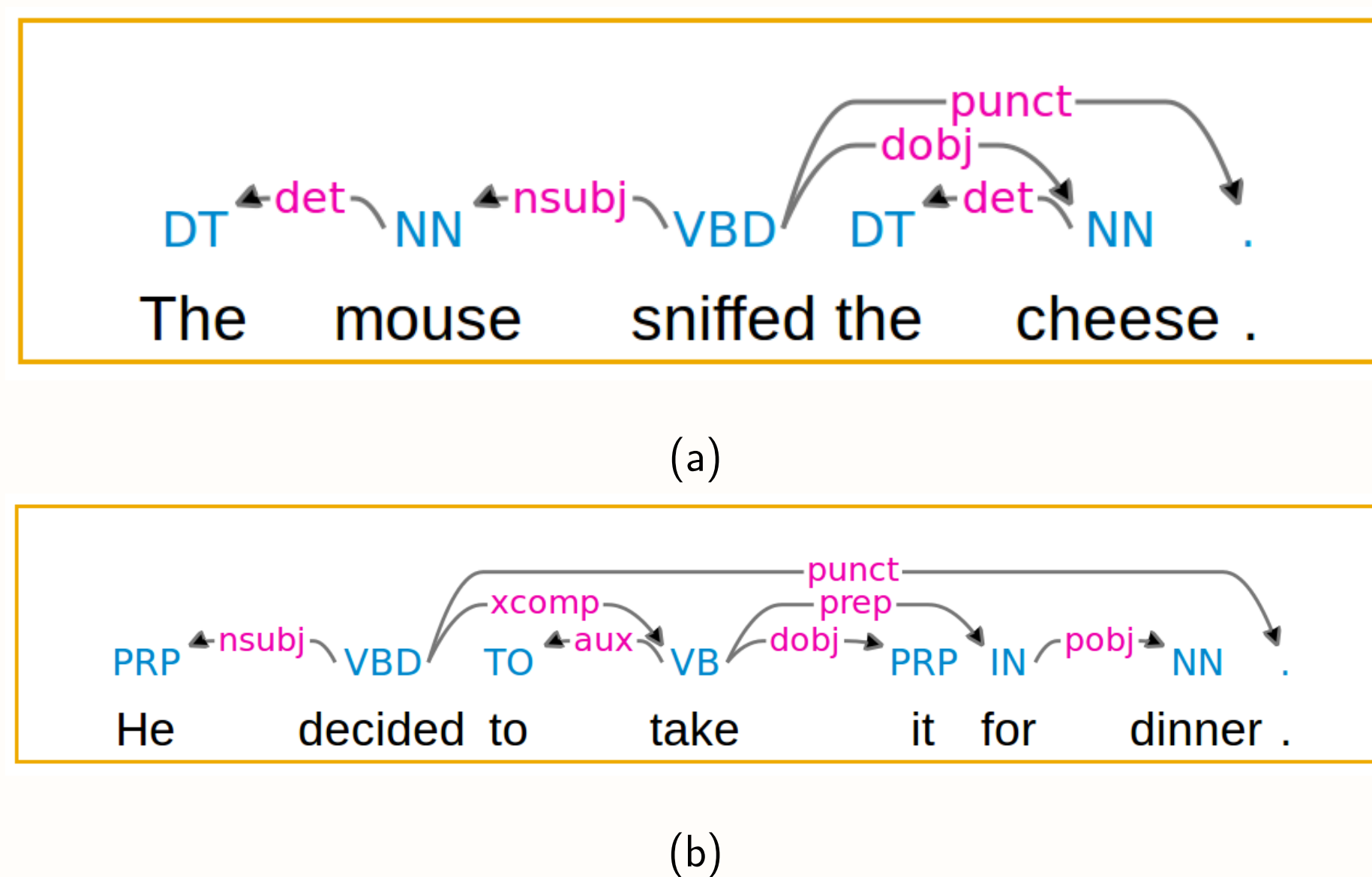


Figure 1 : Example of syntactic parsing (pink) and part-of-speech (POS) tagging (light blue) of two (consecutive) sentences. Example parsed using SEMAFOR [3].

However, **Semantic Parsing** – extracting meaning from text – is a quite new, far away from solved, and challenging problem (figure 2).

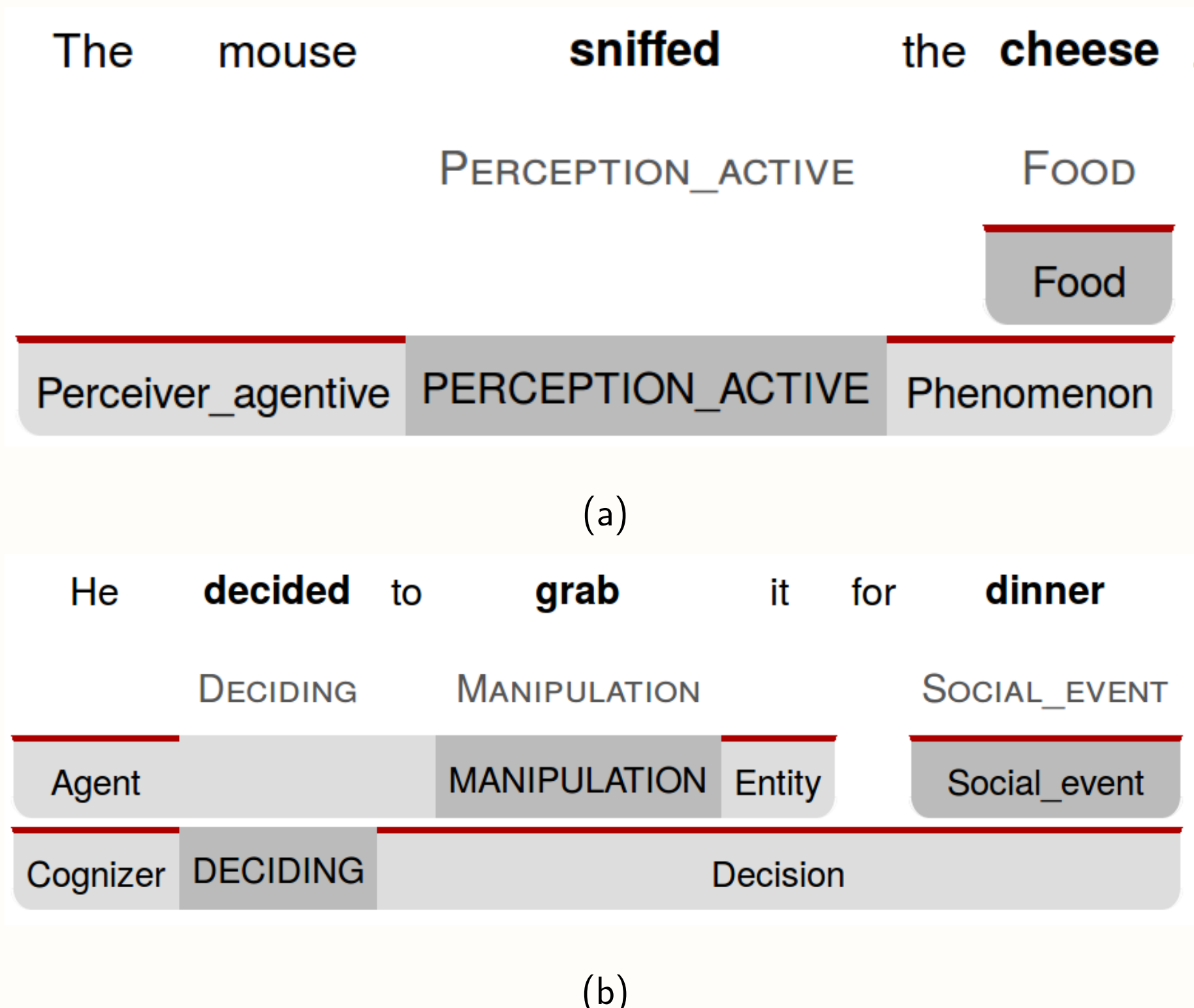


Figure 2 : Semantic parsing of sentences in figure 1. Example parsed using SEMAFOR [3].

Motivation

Interest in diverse fields of NLP has exploded over last years due to many reasons, such as the (every day) bigger amount of textual information available and shared daily through social networks.

Research in **Semantic Parsing** will lead us

- to better **search engines**, looking for meaning instead of only word matches;
- better **Information Retrieval** (IR) from documents;
- or better **Human Computer Interaction** (HCI), etc.

Eventually, semantic parsing would lead us to an **intelligent machine**, being thus an essential step towards **Artificial Intelligence**.

Objectives

Contribute to the research in the Semantic Parsing world following the work done in [5] and [2]. There is still no NLP system able to deal with semantic information in a general free-context manner, and therefore boosting our knowledge from **Natural Language Processing** to **Natural Language Understanding** (NLU), which is our ultimate goal, see figure 3.

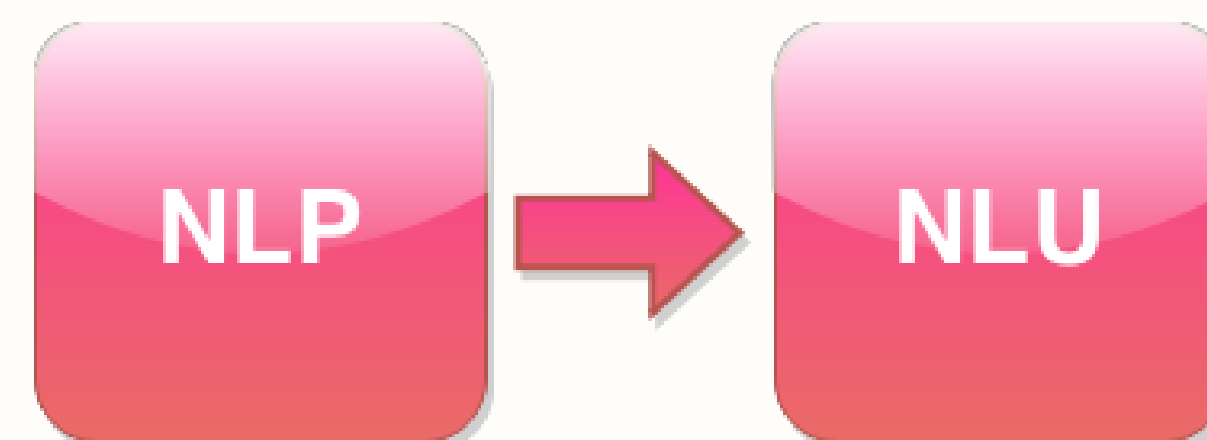


Figure 3 : From Natural Language Processing (NLP) to Natural Language Understanding (NLU).

Research Plan

We attempt to contribute incrementally to computing semantics by publishing our results as articles into international conferences and journals where they will be objectively valued by experts in the field of semantics.

Current State

We are currently working on:

- **FrameNet parser**, we are parsing FrameNet to make it suitable for our own purposes (see Resources subsection).
- **Overview paper**. There are many different approaches to Semantic Parsing. Comparing them would help us and new researchers to get a more general perspective of the state of the art of semantic computing. This paper is currently being written.

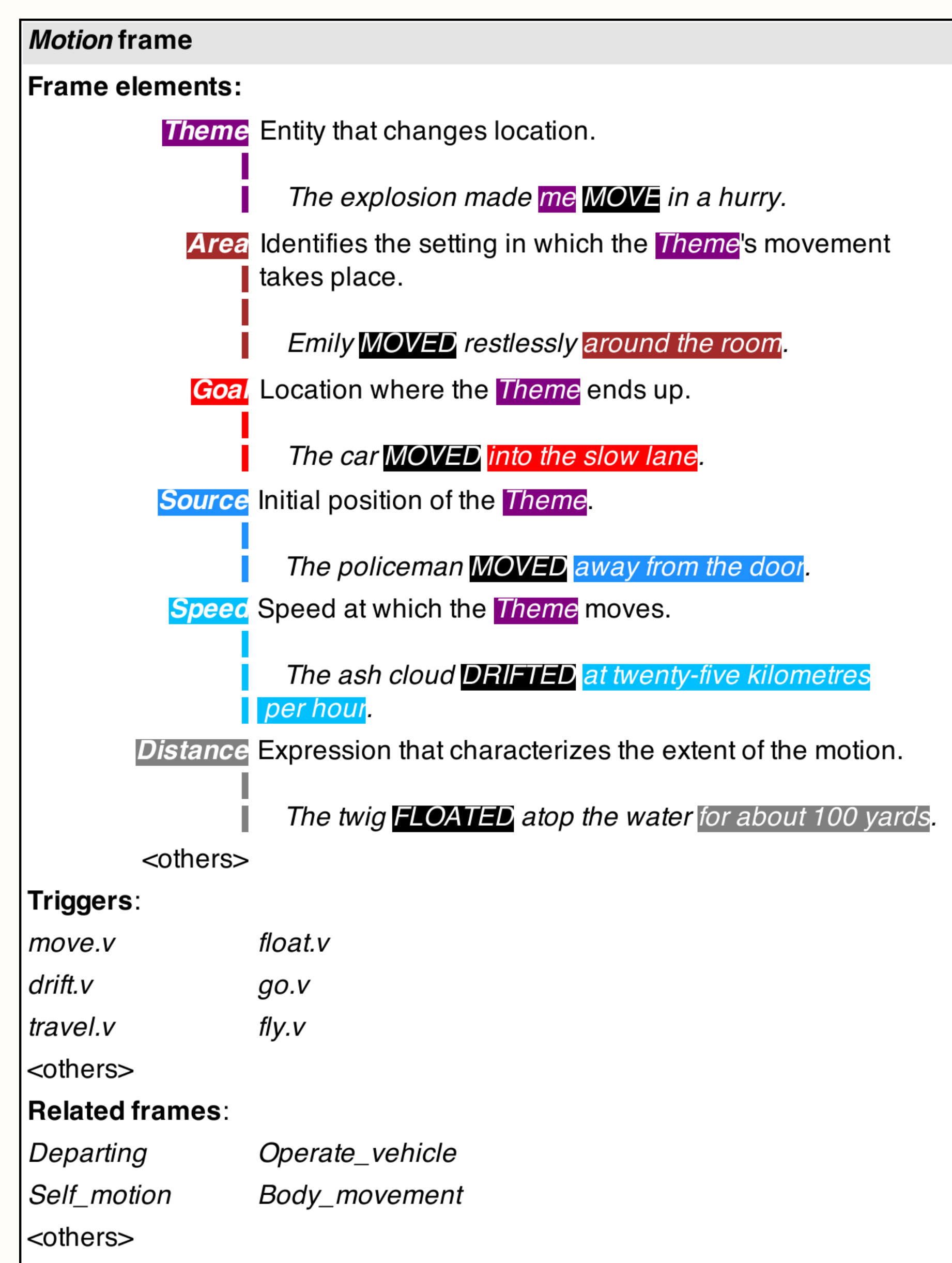


Figure 4 : Extract of the Motion frame, showing some of its frame elements, triggers and related frames.

Resources

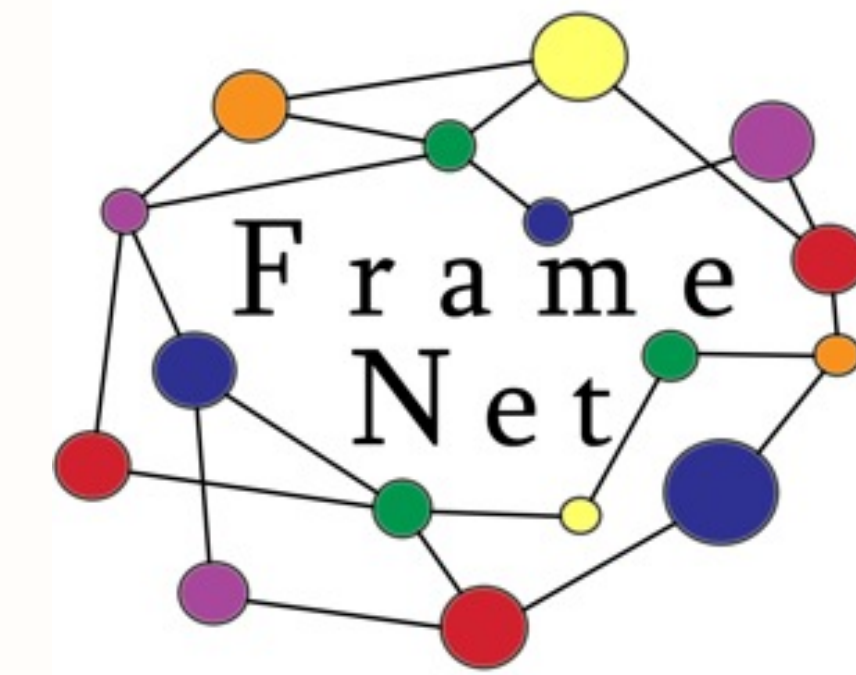


Figure 5 : FrameNet logo.

FrameNet [1] (figure 5) is an ontology of related concepts, called frames, based on Fillmore's framing theory [4].

A frame represents a concrete interpretation of reality and has attached to it several components and triggers. For example, if we are in a motion scenario (*something is moving*) we expect to find some (*frame*) elements: the one in motion (*Theme*), or initial (*Source*) and final (*Goal*) points of the movement, see figure 4.

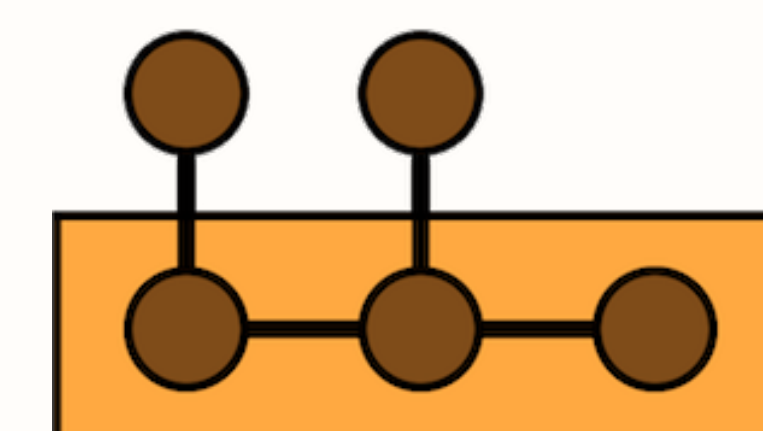


Figure 6 : Factorie logo.

Factorie (figure 6) is a toolkit for probabilistic modeling. It is implemented as a software library written in Scala. It is intended to use to train classifiers on text able to extract FrameNet relations.

Other resources are intended to be used in the future, like **WordNet** and **VerbNet**.

Planning

- **Continue reading literature** about the topic.
- **Finishing and publishing** our overview paper.
- **Start building our Semantic Parsing system** and have a (pre-)alpha, version during next months.
- **Assist to conferences** about these or related topics to share ideas with other researchers and look for support whenever possible.

References

- [1] C. F. Baker, C. J. Fillmore, and J. B. Lowe. The berkeley framenet project. In *Proceedings of the 17th International Conference on Computational Linguistics - Volume 1, COLING '98*, pages 86–90, Stroudsburg, PA, USA, 1998. Association for Computational Linguistics.
- [2] D. Das. *Semi-Supervised and Latent-Variable Models of Natural Language Semantics*. PhD thesis, Carnegie Mellon University, 2012. <http://repository.cmu.edu/dissertations/342>.
- [3] D. Das, D. Chen, A. F. T. Martins, N. Schneider, and N. A. Smith. Frame-semantic parsing. *Comput. Linguist.*, 40(1):9–56, Mar. 2014.
- [4] C. J. Fillmore. Frame semantics and the nature of language. *Annals of the New York Academy of Sciences: Conference on the Origin and Development of Language and Speech*, 280(1):20–32, 1976.
- [5] D. Gildea and D. Jurafsky. Automatic labeling of semantic roles. *Comput. Linguist.*, 28(3):245–288, Sept. 2002.
- [6] A. McCallum, K. Schultz, and S. Singh. FACTORIE: Probabilistic programming via imperatively defined factor graphs. In *Neural Information Processing Systems (NIPS)*, 2009.