

# Semantic Role Labeling: Past, Present and Future

Lluís Màrquez  
TALP Research Center  
Software Department  
Technical University of Catalonia  
lluism@lsi.upc.edu

## 1 Introduction

Semantic Role Labeling (SRL) consists of, given a sentence, detecting basic event structures such as “who” did “what” to “whom”, “when” and “where”. From a linguistic point of view, a key component of the task corresponds to identifying the semantic arguments filling the roles of the sentence predicates. Typical predicate semantic arguments include Agent, Patient, and Instrument, but semantic roles may also be found as adjuncts (e.g., Locative, Temporal, Manner, and Cause). The identification of such event frames holds potential for significant impact in many NLP applications, such as Information Extraction, Question Answering, Summarization and Machine Translation.

Recently, the compilation and manual annotation with semantic roles of several corpora has enabled the development of supervised statistical approaches to SRL, which has become a well-defined task with a substantial body of work and comparative evaluation. Significant advances in many directions have been reported over the last several years, including but not limited to: machine learning algorithms and architectures specialized for the task, feature engineering, inference to force coherent solutions, and system combinations.

However, despite all the efforts and the considerable degree of maturity of the SRL technology, the use of SRL systems in real-world applications has so far been limited and, certainly, below the initial expectations. This fact has to do with the weaknesses and limitations of current systems, which have been highlighted by many of the evaluation exercises and keep unresolved for a few years (e.g., poor generalization across corpora, low scalability and efficiency, knowledge poor features, too high complexity, absolute performance below 90%, etc.).

## 2 Content Overview and Outline

This tutorial has two differentiated parts. In the first one, the state-of-the-art on SRL will be overviewed, including: main techniques applied, existing systems, and lessons learned from the CoNLL and SemEval evaluation exercises. This part will include a critical review of current problems and the identification of the main challenges for the future. The second part is devoted to the lines of research oriented to overcome current limitations. This part will include an analysis of the relation between syntax and SRL, the development of joint systems for integrated syntactic-semantic analysis, generalization across corpora, and engineering of truly semantic features. See the outline below.

1. Introduction
  - Problem definition and properties
  - Importance of SRL
  - Main computational resources and systems available for SRL
2. State-of-the-art SRL systems
  - Architecture
  - Training of different components
  - Feature engineering
3. Empirical evaluation of SRL systems
  - Evaluation exercises at SemEval and CoNLL conferences
  - Main lessons learned
4. Current problems and challenges
5. Keys for future progress
  - Relation to syntax: joint learning of syntactic and semantic dependencies
  - Generalization across domains and text genres
  - Use of semantic knowledge
  - SRL systems in applications
6. Conclusions