

Why Don't Romanians Have a Five O'clock Tea, Nor Halloween, but Have a Kind of Valentines Day?

Corina Forăscu

University A.I. Cuza of Iași, Faculty of Computer Science
Research Institute for Artificial Intelligence, Romanian Academy
16, Gen. Berthelot, Iași – 700483, Romania
corinfor@info.uaic.ro

Abstract. Recently the focus on temporal information in NLP applications has increased. Based on general temporal theories, annotations and standards, the paper presents the steps performed towards obtaining a parallel English-Romanian corpus, with the temporal information marked in both languages. The automatic import from English to Romanian of the TimeML markup has a success rate of 96.53%. The paper analyzes the main situations that appeared during the automatic import: perfect or impossible transfer, transfer with amendments or for the language specific phenomena. This corpus study permits to decide how import techniques can be used on the temporal domain.

1 Introduction

The temporal information is expressed in natural language through:

- Time-denoting temporal expressions – references to a calendar or clock system, expressed by NPs, PPs, or AdvPs, as in *Friday; yesterday; the previous month*.
- Event-denoting temporal expressions – explicit/implicit/vague references to an event; syntactically they are realized through:
 - sentences – more precisely their syntactic head, the main verb, as in *She flew as the first ever co-pilot*.
 - noun phrases, as in *She followed a normal progression within NASA*.
 - adjectives, predicative clauses or prepositional phrases, as in : *Many experts thought was once invincible*.

Recent work in document analysis started focusing on the temporal information in documents, mainly for their use in many practical Natural Language Processing (NLP) applications such as:

- linguistic investigation, lexicon induction, and translation using very large annotated corpora;
- question answering (questions like “when”, “how often” or “how long”);
- information extraction or information retrieval;
- machine translation (translated and normalized temporal references; mappings between different behavior of tenses from language to language);