

# **A Unified Theory of Ellipsis through Contrast**

by

Carolina Soledad Vedovato

A Thesis Submitted to the Graduate  
Faculty of Rensselaer Polytechnic Institute  
in Partial Fulfillment of the  
Requirements for the degree of  
MASTER OF SCIENCE  
Major Subject: COGNITIVE SCIENCE

Approved:

---

Nicholas Cassimatis, Thesis Adviser

Rensselaer Polytechnic Institute  
Troy, New York

April 2012  
(For Graduation May 2012)

## ABSTRACT

The lack of true interpretation in natural language processing tools renders users bored, frustrated and ultimately disenchanted, limiting these tools to be used in extreme situations, such as being stranded, being in a car, or being unable to type. We seek to develop approaches that would allow language to be the main communicative tool between computers and humans, as it is among humans ourselves.

Within natural language processing, ellipsis (present in utterances such as *Mary went to the party and Sarah too*) has been a particularly elusive phenomenon, presenting great difficulties in its interpretation due to the mismatch between utterance and meaning. Until now, the theories used to solve ellipsis in language were usually based on a division into many different types of ellipsis, each with its own solution. As a result, sentences as similar as *Mary went to the party and Sara did too* (a case of VP ellipsis) and *Mary went to the party and Sarah too* (an example of stripping) need to be solved in two different ways. In addition, many theorists consider ellipses to be different according to the type of information that has been omitted, be it syntactic, semantic or pragmatic. Furthermore, some utterances (cases of compounded sloppy and strict readings in particular) lose possible interpretations because of exceedingly restrictive definitions of ellipsis that consider some of the possible combinations of readings to be

ungrammatical (even though there are possible contexts of application for such interpretations).

In order to better explain ellipsis in a simpler and more unified manner, we adopt a theory based on the principle of contrast, and apply these findings to different cases of ellipsis through first-order probabilistic constraints. To this end, we utilize the concept of contrast to identify the omitted elements in any case of ellipsis and anchor them to their most likely identity match through the comparison of elements related to the syntax, semantics and pragmatics of the utterance.

In this manner, we can identify what elements have been elided, and analyze the similarities and differences between the possible candidates in order to find the best possible match to the elided utterance or fragment. This approach is further aided by a deeper understanding of the positive role of ungrammaticality as a closer representation of natural language. This allows us to process cases of ellipsis and strict/sloppy readings usually left behind by other theories in the very same manner we deal with the most common types of ellipsis.