Abstract

This dissertation is a quantitative analysis of constituent order (i.e. the order of subject, verb and object) in contemporary (post-2000) Maltese, a Semitic language descended from North African varieties of Arabic, spoken primarily in the Malta archipelago and Australia. The analysis is based on data collected in two corpora: a general corpus and a syntactically annotated corpus (dependency treebank); the compilation and description of the treebank is the secondary aim of this work.

The dissertation comprises 8 chapters divided into two parts: the first three chapters of which provide a conceptual foundation (chapter 1), a review of existing major approaches to the study of constituent order (chapter 2) and a review of previous works on Maltese constituent order (chapter 3). Using these as the background, chapter 4 then sets the research questions and methodology. The remaining three chapters form the core of the dissertation: chapter 5 describes the composition and enrichment of the general corpus of Maltese. Chapter 6 then provides a thorough description of the Maltese treebank and the annotation decisions, thus in effect assembling a sketch of Maltese syntax. Chapter 7 then provides the actual quantitative analysis of constituent order in Maltese based on the treebank, arriving at the conclusion that the dominant order is SVO or SV/VO and making a detailed case for rejecting previous descriptions of Maltese constituent order as "discourse-configurational", "topic-prominent" and "free". The final chapter summarizes the findings and lays out a number of avenues for further research into the topic.

Key words: Maltese, computational linguistics, syntax, dependency syntax, treebank, constituent order