Theodorus Leman Franciscus Fransen

Past, present and future: Computational approaches to mapping historical Irish cognate verb forms

Abstract

This thesis investigates how computational methods can be employed to enhance our understanding of the significant developments in the verbal system between Old Irish (c. 8th–9th centuries A.D.) and Modern Irish (13th century onwards). The most fundamental changes happen in the course of the Middle Irish period (c. 10th–12th centuries), resulting in a much-simplified verbal system in the Modern Irish period, where verb forms have often changed beyond recognition relative to their Old Irish predecessors. At the same time, however, one is faced with a lack of lexical resources for the intermediate stages between Old Irish and contemporary Modern Irish, which impedes a systematic investigation of the extensive linguistic changes occurring between these language periods. The main contribution of the present work is the creation of a morphological Finite-State Transducer (FST) for Old Irish verbs, successfully implemented in the finite-state tool `foma` (Hulden 2009). An automatic morphological parser for Old Irish verbs is deemed invaluable for comprehensively and systematically interlinking diachronic variants. This thesis documents the challenges encountered and choices made relative to the finite-state morphological implementation, in the context of an extremely complex interplay between morphology and phonology. The current work puts forward a computationally workable definition of a verb stem as a solution for this complexity problem. It was decided to meaningfully limit the scope of the project by focusing on Old Irish weak verbs, whose inflection patterns, compared to the strong verb classes, are more transparent and predictable. Moreover, weak verb inflection becomes the dominant type in Middle Irish. After having incorporated into the FST stems for 27 weak verbs found in the text *Táin Bó Fraích* (Meid 1974, Meid et al. 2015), as well as a limited amount of additional frequent items, a recognition rate of 9.6% was obtained for word types, which is consistent across four other Old Irish texts (on average 10%). The current work concludes with a roadmap for the future, which involves a bidirectional diachronic mapping framework for Irish verbs. This framework consists of two key resources: the Old Irish morphological FST developed as part of this thesis, and an already available FST for standardised contemporary Modern Irish (Uí Dhonnchadha & van Genabith 2006).