Chron Hib
Chronologicon Hibernicum
ERC H2020 no. 647351

Ollscoil Mhá Nuad
Ollscoil na hÉireann Má Nuad

Computational linguistics and the dating of early Irish Texts

The *Chronologicon Hibernicum* project is hosting a series of three thirty-minute public lectures on 15 December, 2016.

These lectures will explore computational methods for linguistic research, especially as they apply to the study of Old and Middle Irish. Recent projects have made an increasing number of digital humanities resources available to the researcher. These include linguistically-parsed corpora, such as the Old Irish glosses databases (*Milan Glosses*, http://www.univie.ac.at/indogermanistik/milan_glosses.htm, Griffith 2013; *Priscian Glosses*, http://www.univie.ac.at/indogermanistik/priscian/, Bauer 2015) and the *Parsed Old and Middle Irish Corpus* (https://www.dias.ie/celt/celt-publications-2/celt-the-parsed-old-and-middle-irish-corpus-pomic/, Lash 2014) as well as a number of text repositories, such as *CELT* (https://www.ucc.ie/celt/) and *Thesaurus Linguae Hibernicae* (http://www.ucd.ie/tlh/).

Harnessing the potential of these resources and developing new publicly available corpora in order to help date Old and Middle Irish texts and to gain deeper insights into the development of the phonology, morphology, syntax, and lexicon of the Irish language is a major goal of the *Chronologicon Hibernicum* project.

The lectures will **start at 15:00** and run for approximately two and half hours with a coffee/tea break between each lecture.

Location: Staff Development Seminar Room, John Hume Building, Maynooth University (Take the lift to the third floor. Exit right and follow the corridor to the end. The room is in the far left-hand corner.)

Programme:

David Stifter (Maynooth University) **Introduction** - 15:00 - 15:05

Marius Jøhndal (University of Oslo) **"Building and using online corpora for (historical) linguistic research"** - 15:05 - 15:50.

"This talk will look at how to build and use corpora for research in historical linguistics. In particular, I will look at the experiences of the PROIEL Treebank project, which compiled a corpus of syntactically annotated Indo-European texts, and the degree to which computational approaches can be applied when building corpora of historical languages."

Aaron Griffith (Utrecht University) **"Pre-verbal *ceta* 'first' in the glosses (and some thoughts of the origin of the *notae augentes*)"** - 15:55 - 16:40.

"This talk consists of two parts. In the first, I examine the distribution of pre-verbal *ceta* 'first' < **kintu*- in the glosses, in order to determine the developments. Following that, I discuss the notae augentes, the paradigm of which, under any interpretation, is of heterogeneous origin. I attempt to offer an explanation of this heterogeneity."

Gregory Toner (Queen's University, Befast) **"Machine learning and the dating of Medieval Irish Texts"** - 16:45 - 17:30.

This paper will describe the use of computer-based methods for the dating of the medieval Irish corpus adopted by the LexiChron project at Queen's University Belfast and analyse some of the preliminary results. The method use various sets of annals to train a classifier to construct a picture of changes in language in the annals and has achieved a 75% success rates in dating a test set of annals to within +/- 25 years. The paper will explore the potential for transferring this method to non-annalistic texts both for absolute and relative dating.