

As is the case with many languages, research into code-switching in Modern Irish has, until recently, mainly been focused on the spoken language. This paper reports on the annotation of (English) code-switching in a corpus of 1496 Irish tweets and provides a computational analysis of the nature of code-switching amongst Irish-speaking Twitter users, with a view to providing a basis for future linguistic and socio-linguistic studies. Online user-generated content (UGC) is less restrictive than traditional written text, allowing for code-switching, and as such, provides a new platform for text-based research in this field of study. Analysis of minority language UGC in particular provides much insight into the evolution of these languages in the digital age.

1 Introduction

User-generated content (UGC) provides an insight into the use of language in an informal setting in a way that previously was not possible. That is to say that in the pre-internet era (where most published content was curated and edited), text that was available for analysis was not necessarily reflective of everyday language use. In some bilingual environments, the overwhelming dominance of a majority language can sometimes restrict and discourage the natural use of a minority language. Additionally, various interesting linguistic phenomena occur online that may be frowned upon in more formal settings. The present paper aims to investigate one such phenomenon among Irish-speaking users of the micro-blogging platform Twitter. This article is licensed under a Creative Commons 4.0 licence, no derivative works, attribution, CC-BY-ND.c

.2019 The authors