Native Language Identification for Forensic Authorship Analysis

Dr Ria Perkins completed her PhD in 2013 at Aston University. Her thesis focused on **interlingual identifiers** of L1 Persian speakers blogging in English. Since finishing her PhD Ria is moving on to research linguistic structures in English and Swedish language texts of online extremism and radicalisation. She is now working at Aston University as a Teaching Associate.

Context

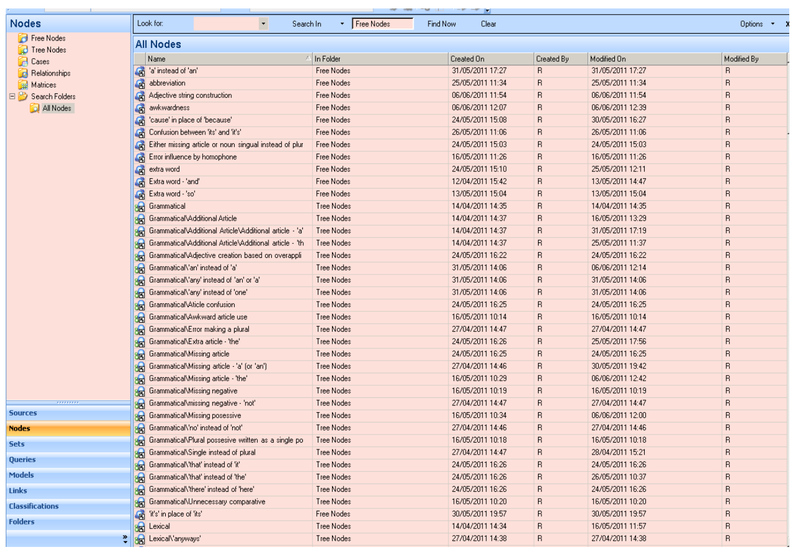
Grounded in Forensic Authorship analysis, this research focused on online blogs from weblogistan to identify interlingual features of L1 Persian speakers writing in English and to develop an implementable model for forensic authorship cases. The potential importance of Native Language Identification (NLID) can be witnessed through the prevalence of multilingualism as well as documented cases such as those of Kniffka (1996 & 2000) and Hubbard (1996). Despite potential practical applications research into NLID from a forensic perspective is limited ; a void which this research seeks to fill.

Data and Analysis

3 sub-studies involving different corpora:

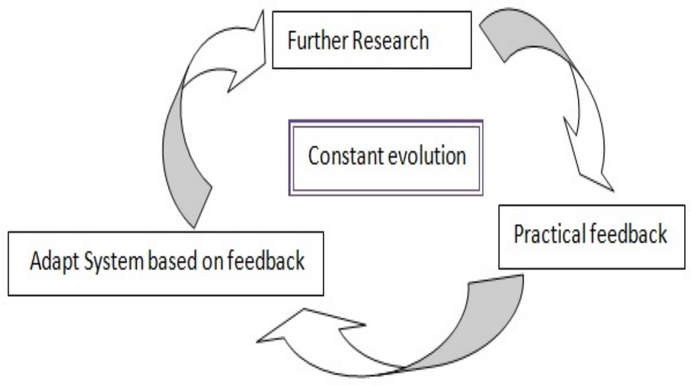
- L1 Persian and L1 English Weblogs
- Other languages (Azeri and Pashto)
- Disguise language

A coding system was developed to account for all the features. The data were then coded using NVivo. Logistic regression was used to determine which features had the higher discriminatory power. The features were then refined to discover the optimum models.



Findings

This research demonstrated that interlingual features in L2 writing can be used to indicate an author's L1. It developed a new methodology for NLID and distinguished which features best indicate authorship by an L1 Persian speaker . The features identified were determined to have a statistically high level of reliability at determining group membership. An implementable model was constructed to help determine if an anonymous text was written by an L1 Persian speaker, as well as distinguishing from close languages and author's attempting to disguise their L1. NLID and the model created have a strong potential to be a valuable tool for forensic authorship analysis and may prove useful for criminal and intelligence investigations. It is intended as part wider research and there is clear potential for further research.



Aston University
Birmingham

Centre for Forensic Linguistics