Language and Gender Author Cohort Analysis of E-mail for Computer Forensics

By

Olivier De Vel, Malcolm Corney, Alison Anderson, George Mohay

Presented At

The Digital Forensic Research Conference

DFRWS 2002 USA Syracuse, NY (Aug 6th - 9th)

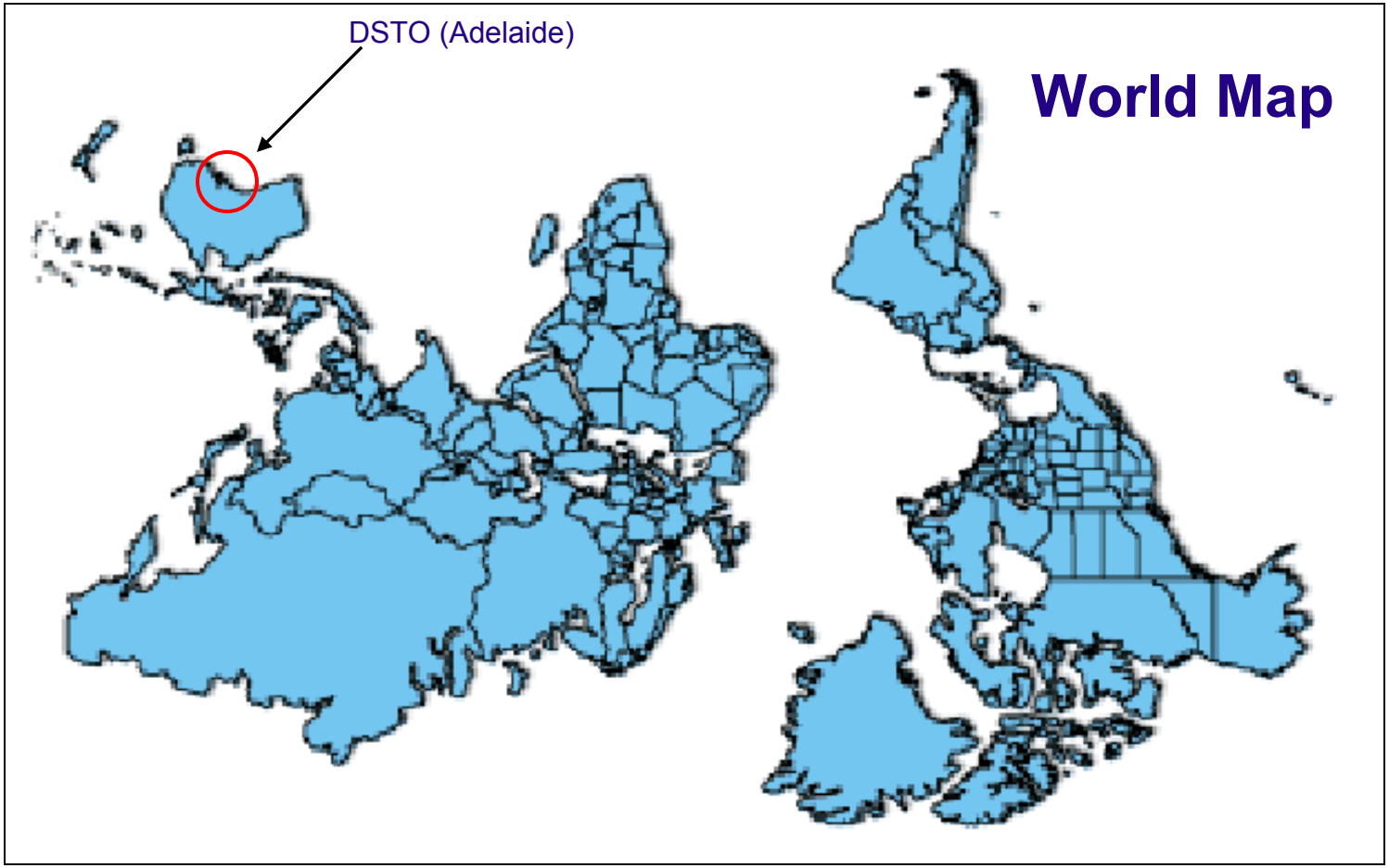
DFRWS is dedicated to the sharing of knowledge and ideas about digital forensics research. Ever since it organized the first open workshop devoted to digital forensics in 2001, DFRWS continues to bring academics and practitioners together in an informal environment. As a non-profit, volunteer organization, DFRWS sponsors technical working groups, annual conferences and challenges to help drive the direction of research and development.

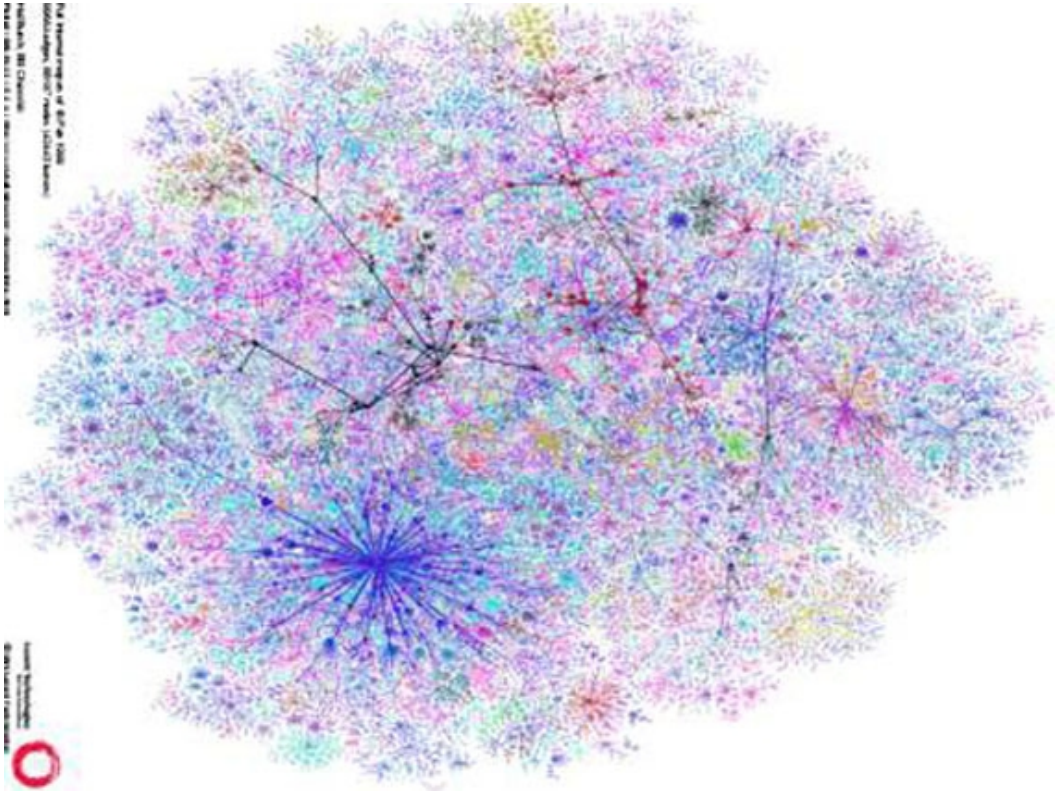
http:/dfrws.org

Language and Gender Author Cohort Analysis of E-mail for Computer Forensics

Olivier de Vel
Computer Forensics

DSTO, Australia


DSTO (Adelaide)
World Map



Presentation:

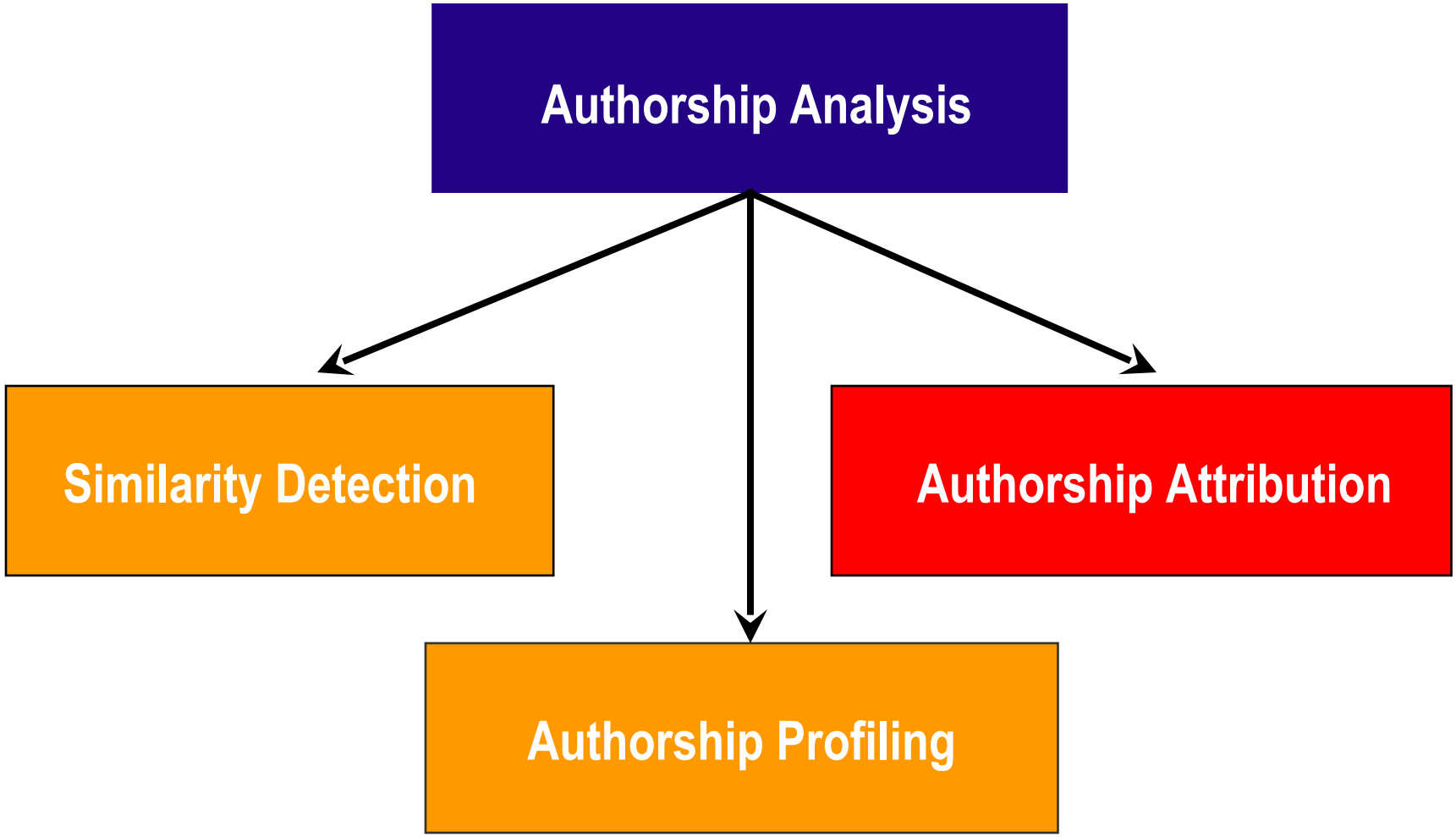
Computer forensics and e-mail authorship analysis

Previous work in authorship attribution

Experimental methodology

Results and Conclusion

Authorship Analysis: Sub-Areas

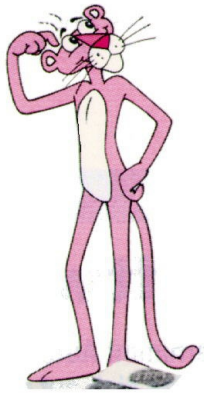


Objective of E-mail Authorship Analysis

Develop algorithms for analysing the style and content of an e-mail message for the purpose of categorising

- its author, or
- its author's cohort type

E-mail authorship analysis is NOT about:



- e-mail document filtering
- e-mail document filing
- e-mail text categorisation
- e-mail topic detection/tracking

Presentation:

Computer forensics and e-mail authorship analysis

Previous work in authorship attribution

Experimental methodology

Results and Conclusion

Previous Work in Authorship Attribution

- Shakespeare's works & Federalist papers
- Forensic linguistics
- Code authorship

Previous Work in Authorship Attribution: Issues

Conclusions (so far!):

- a large number of stylometric features (>1000)
- no definite subset of discriminatory stylometric features
- no consensus on methodology
- Questionable analysis
- Hard problem!

Previous Work in Authorship Attribution

Previous work limited to (c.f. e-mails):

- Large sections of text
- Formal text and non-interactive
- Relatively large number of training examples
- Relatively homogeneous style

Previous Work in Authorship analysis: E-mails

E-mail authorship attribution:

- We have investigated the effect of several parameters (eg., text size, number of documents per author) [2000, 2001].
- Thomson *et al* have investigated gender-preferential language styles [2001].

Presentation:

Computer forensics and e-mail authorship analysis

Previous work in authorship attribution

Experimental methodology

Results and Conclusion

Experimental Methodology: E-mail Corpus

Difficulty in obtaining e-mail corpus:

- **privacy and ethical concerns,**
- **large % of noise (eg, cross-postings, off-the-topic spam, empty body with attachments),**
- **some difficulty in verifying author cohort class,**
- **non-orthogonality of topics and cohorts.**

Experimental Methodology: E-mail Corpus

Two author cohort experiments (gender and language)

- Two sub-corpuses derived from an E-mail corpus:
  - M/F: 325 authors, 4369 e-mails
  - EFL/ESL: 522 authors, 4932 e-mails

Experimental Methodology: E-mail Corpus

Attributes/features used:

- 183 style markers that are known to have reduced content bias (incl. function words and word freq. distribution),
- 28 e-mail structural attributes,
- 11 gender-preferential language attributes

Experimental Methodology: Classifier

- SVM^{light} as the (two-way) classifier,
- Obtain two-way categorisation matrix for each author category, using 10-fold cross-validation sampling,
- Calculate per-author category performance statistics – precision, recall and F_1 statistics.

Presentation:

Computer forensics and e-mail authorship analysis

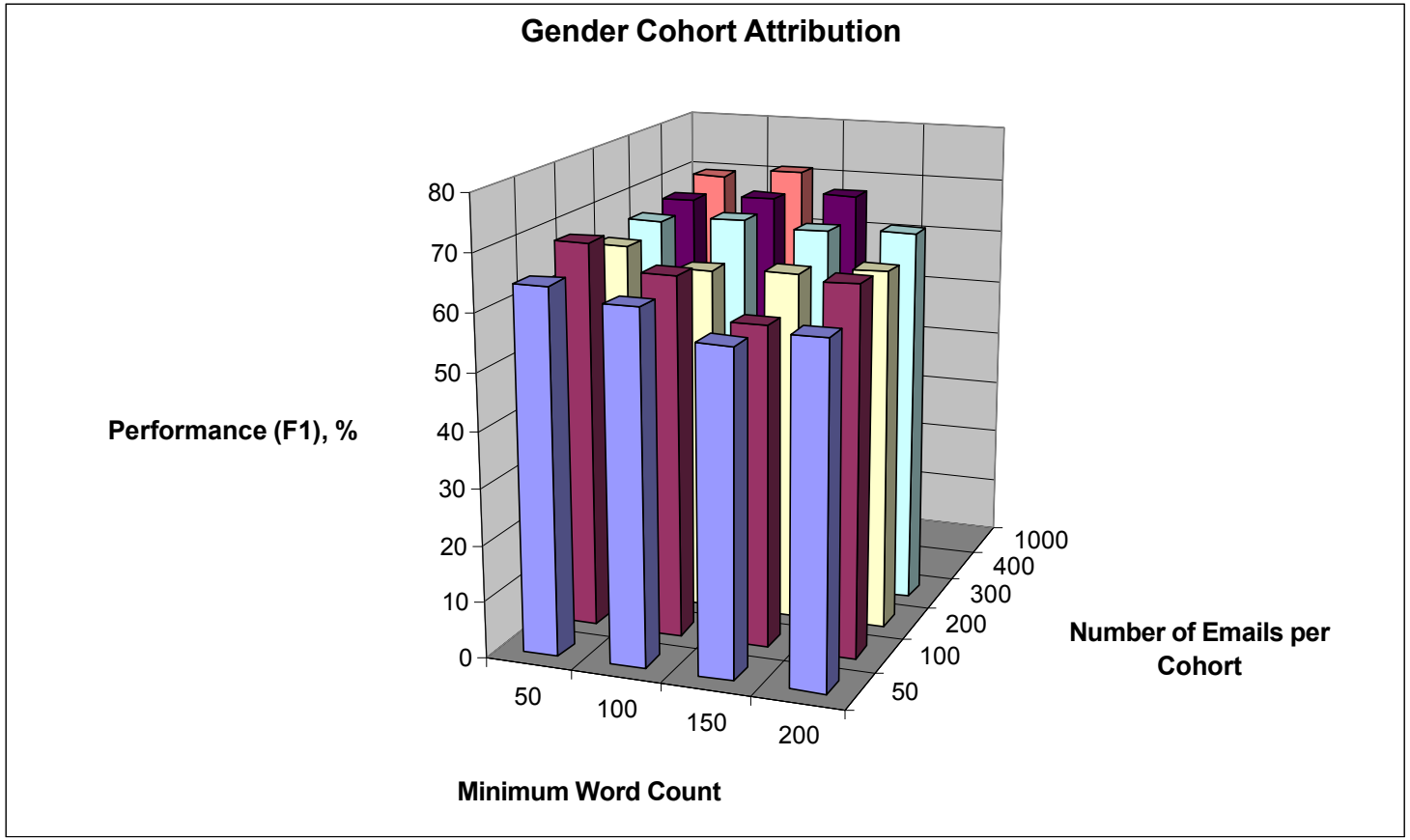
Previous work in authorship attribution

Experimental methodology

Results and Conclusion

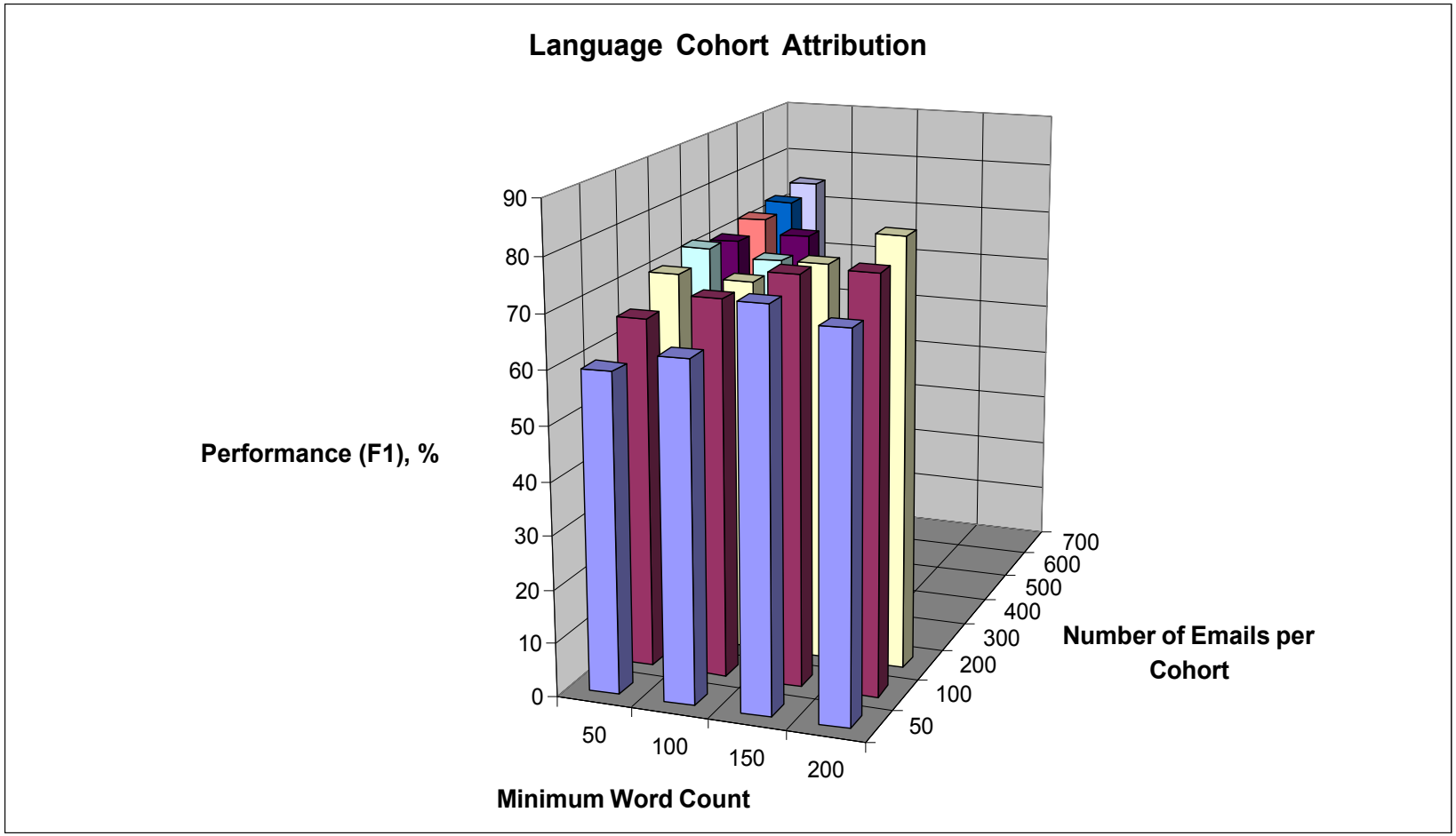
Results: Classification performance

M/F Author Cohort Attribution:



Results: Classification performance

EFL/ESL Author Cohort Attribution:



Conclusions:

- Promising author cohort categorisation results.
- Further experiments:
  - with extended and specific set of cohort-preferential attributes,
  - more within-cohort diversity,
  - subset feature selection (particularly function words).
- Extend to other forms of computer-mediated communications (chat rooms etc.)

Questions ?