

Thesaurus of British Surnames

Minutes of Meeting of Steering Group 9/5/02 at SoG

Present:

Ian Galbraith, Peter Christian, Kevin Schürer, Steve Archer, John Dawson, John Challis

1. Test Data

1.1 IG suggested providing Boyd's Marriage Index (BMI) surname variants data. The value of this was discussed. The surnames in BMI are often partially standardised, and not transcriptions of what were in the original documents, but the variants which are identified could be used to compare against the results of the ToBS work. IG will provide a list of BMI surname groups. (NB This will cover only counties for which data has been captured to date.)

IG

1.2 PC commented upon the interchangeability of the initial consonant: he felt that this was overstated, and that his research showed very few surname variants where the initial consonant varied. (The "equivalence" of, eg C and K are among the few common examples.) IG commented upon the substitution of "C" for "G", eg Culbraith & Galbraith.

1.3 1881 census data. Felt to be a big enough sample for algorithm tests. Groups such as BMI can be used for checking. KS will send JC a test 1881 dataset.

KS

1.4 Hanks & Hodges, Oxford Dictionary of British Surname: KS will seek permission to use this, but this may not be received. Estimated 4,000 entries, c.8-10K total names.

KS

1.5 Scottish Old Parish Registers (OPRs): IG will prepare surname lists from the OPRs, giving surname, date and county. (NB Will need permission from LDS and from GROS.) Copies to PC & JC.

IG

1.6 NBI: SA will provide a list of NBI surnames (175K names).

SA

1.7 GOONS: PC will seek lists of GOONS surnames from John Hanson.

PC

2. TOBS Web Site

KS has arranged for certain documents to be passworded, for access only by SG members. IG will distribute the password to other SG members. (NB This does not seem to have been done, and the minutes of the last meeting are not yet on the site.)

IG

3. Measurement

PC will contact Brian Randall regarding use of the test procedures which were used by his student for measuring effectiveness of grouping algorithms.

PC

4. John Challis work

JC described the work he was doing and the multi-algorithm approach he used. He would put a description on his site (www.imagepartners.co.uk/Thesaurus).

JC

5. Literature search

5.1 A full literature search needed to be done, including a check on the intellectual property aspects of specific algorithms. This work should be put into a Marc Fitch application.

5.2 IG suggested that a request be posted on the mail list for volunteers to do a literature search. There is a phonology mailing list.

6. Funding

6.1 Marc Fitch Fund should be an agenda item for next ToBS meeting to discuss what we might propose as a project for funding. KS to advise date of next Marc Fitch meeting.

KS

6.2 IG to chase BL re contacting Keith Cole at MIMAS.

IG

7. Next meeting

Thursday, 11 July 2002, 1400, at Society of Genealogists, 9 Dallington Street, London.