The Potential of the Historical Thesaurus of English

Christian Kay
University of Glasgow

Historical Text-Mining Workshop, Lancaster 20/7/06
Historical Thesaurus of English

- Oxford English Dictionary (OED, second edition)
- Dictionaries of Old English
- 650,000 meanings in 26 major semantic fields.

http://www.arts.gla.ac.uk/sesll/englang/thesaur/homepage.htm
State of Play

- MS Access database
- Apache HTTP
- MySQL
- PHP (Hypertext Preprocessor)
- *Thesaurus of Old English*: TOE-online

- [http://libra.englang.arts.gla.ac.uk/oethesaurus/](http://libra.englang.arts.gla.ac.uk/oethesaurus/)
Basic searches

http://leo.englang.arts.gla.ac.uk/historicalthesaurus/

Synonyms in category 03.01.01.03.03. ./ 02. . . . - n - (.grandfather)

<table>
<thead>
<tr>
<th>Main heading</th>
<th>Word</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Grandparent</td>
<td>ealda fæder</td>
<td>n</td>
<td>OE</td>
<td></td>
</tr>
<tr>
<td>Grandparent</td>
<td>ieldra fæder</td>
<td>n</td>
<td>OE</td>
<td></td>
</tr>
<tr>
<td>Grandparent</td>
<td>eldfather</td>
<td>n</td>
<td>OE -</td>
<td>1460</td>
</tr>
<tr>
<td></td>
<td>&lt; ealdefæder</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Grandparent</td>
<td>grandsire</td>
<td>n</td>
<td>c 1290 -</td>
<td>1876 ai&amp;dl</td>
</tr>
<tr>
<td>Grandparent</td>
<td>aiel</td>
<td>n</td>
<td>1377 -</td>
<td>1502</td>
</tr>
<tr>
<td>Grandparent</td>
<td>belsire</td>
<td>n</td>
<td>1377 - a</td>
<td>1631</td>
</tr>
<tr>
<td>Grandparent</td>
<td>grandfather</td>
<td>n</td>
<td>1424 -</td>
<td></td>
</tr>
</tbody>
</table>
Lexical Searches for the Arts and Humanities

(AHRC-ICT Resource Development Project AR112456)

- to demonstrate how a project developed largely to meet the needs of one field (English Historical Linguistics) can have applications in other disciplines.

- to demonstrate how electronic resources can enhance, and indeed promote, research in the Arts and Humanities.
Please select the type of search you wish to do.

**Browse**
This will search the whole database and display the section names(s) where a target word is located. Then proceed to browse headings, words and dates under concept. This is especially useful where a word occurs in more than one section.

**Synonym search**
As above and leads on to display synonyms under the chosen heading.

**Label**
List words and dates with selected style & status labels.

**Affix**
Lists words with specified affixes, the dates and the headings under which they occur.

**Part of speech**
Lists words restricted by part of speech, the dates and the headings under which they occur.

**Dates**
Choose type of date search then follow on to choose section for date search.
Currently available searches

- **Browse**
  Search the whole database and display the section names(s) where a target word is located. Then proceed to browse headings, words and dates under concept.

- **Synonym search**
  As above and leads on to display synonyms under the chosen heading.

- **Label**
  List words and dates with selected style & status labels.

- **Affix**
  Lists words with specified affixes, the dates and the headings under which they occur.

- **Part of speech**
  Lists words restricted by part of speech, the dates and the headings under which they occur.

- **Dates**
  Choose type of date search then follow on to choose section for date search.
Computational potential

- Detailed and logical taxonomy
  - Clothing ⇔ A hat with a bobble
  - Clothing: 8751 lexical entries; 2212 headings

- Text mining using contemporary words

- Probability-based methods disambiguating polysemous word forms
  - Was *gout* discussed in this 16th century medical text? Are there other books on the subject?
<table>
<thead>
<tr>
<th>Main heading</th>
<th>Word</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Disorders of joints</td>
<td>fotadl</td>
<td>n</td>
<td>OE</td>
</tr>
<tr>
<td>Disorders of joints</td>
<td>fotcoþu</td>
<td>n</td>
<td>OE</td>
</tr>
<tr>
<td>Disorders of joints</td>
<td>fotwærc</td>
<td>n</td>
<td>OE</td>
</tr>
<tr>
<td>Disorders of joints</td>
<td>liþa</td>
<td>n</td>
<td>OE</td>
</tr>
<tr>
<td>Disorders of joints</td>
<td>drop &lt; dropa</td>
<td>n</td>
<td>OE + 1559 (2)</td>
</tr>
<tr>
<td>Disorders of joints</td>
<td>podagre</td>
<td>n</td>
<td>c 1290 - 1578</td>
</tr>
<tr>
<td>Disorders of joints</td>
<td>gout</td>
<td>n</td>
<td>c 1290</td>
</tr>
<tr>
<td>Disorders of joints</td>
<td>podagra</td>
<td>n</td>
<td>1398</td>
</tr>
<tr>
<td>Disorders of joints</td>
<td>arthritic</td>
<td>n</td>
<td>1486</td>
</tr>
<tr>
<td>Disorders of joints</td>
<td>arthritis</td>
<td>n</td>
<td>1544</td>
</tr>
<tr>
<td>Disorders of joints</td>
<td>joint-sickness</td>
<td>n</td>
<td>1545 + 1684</td>
</tr>
<tr>
<td>Disorders of joints</td>
<td>leaping gout</td>
<td>n</td>
<td>1562</td>
</tr>
<tr>
<td>Disorders of joints</td>
<td>goutiness</td>
<td>n</td>
<td>1632 + 1820</td>
</tr>
<tr>
<td>Disorders of joints</td>
<td>wind-gout</td>
<td>n</td>
<td>1662</td>
</tr>
</tbody>
</table>
Contact

Prof. Christian Kay
Dept. of English Language
University of Glasgow
12 University Gardens
Glasgow G12 8QQ

http://www.arts.gla.ac.uk/sesll/englang/thesaur/homepage.htm
Word lovers will be interested in the Daily Telegraph's article about the Historical Thesaurus of the Oxford English Dictionary, which was published yesterday. At nearly 4,500 pages and costing £250, the HTE won't be on many people's Christmas lists, but