<u>Project Update - May 2008</u>

For logistical and developmental reasons, the project has been granted a time extension, whilst still operating within the original funding. The material is currently at an advanced stage of development on an internal website.

Research Associate: Dr Peter Stokes' contracted two years as Research Associate for the LangScape project came to an end in October 2007. The project was exceptionally fortunate to have had the benefit of Peter's expertise - in palaeography, medieval Latin and other scholarly abilities needed over the two years, but particularly for the ease with which he could synthesise these with his computing expertise, communicating ideas and designing and developing complex programs as the project needed them. It was truly a blessing and a pleasure to have had Peter as collaborator on this project.

Update: The **Language of Landscape: Reading the Anglo-Saxon Countryside Project** has, to date, achieved the following goals:

- The electronic Corpus of all known Anglo-Saxon boundary clauses (in both English and Latin) has been completed. These represent all the boundary clauses in charters cited in the Electronic Sawyer (http://www.esawyer.org.uk/content/charter/index.html), plus some hitherto uncited texts and a handful of boundaries in post-Conquest charters. They have all been checked and corrected against their manuscript sources and against the British Academy Anglo-Saxon Charters editions published to date, plus other key editions. These form our 'TopTexts'.

- Virtually all the 'variant versions' of these texts that are cited in the Electronic Sawyer have been transcribed from manuscript and, where relevant, checked against printed editions as above.

- The metadata (manuscript date, hidage figures, archival provenance, grant details etc.) has been fully revised and augmented from the updated Electronic Sawyer, and comprises over one hundred data fields; the structure of the tables allows ongoing revision. In turn, emendations and additions noted during the course of the project have been fed back into the Electronic Sawyer.

- Browsing and search facilities are at an advanced stage of development; these will allow the user to locate the boundary clauses on a map of England, or to browse, search and order boundaries and their metadata according to criteria such as manuscript source, grant date, archive provenance, hidage, etc. This will soon be available in a Beta (experimental) form.

- A system of extensive XML mark-up has been developed to enable each text to be displayed on the website in three 'views' - a semi-diplomatic view, a partially edited view and a lemmatized and 'translated' view, and to allow complex searches on the textual material.

- To this end every text (both top and variant versions) has been broken down into semantically discrete units for lemmatization, whilst preserving the encoded data with respect to the spacing in the manuscript source. An alphabetic list of all these discrete word forms within the material (both English and Latin) was then compiled.

- Based on this list, a comprehensive Glossary database was created using an electronic list of headwords kindly supplied by the Vocabulary of English Place-Names Project (http://www.nottingham.ac.uk/english/ins/about.htm), together with major printed onomastic dictionaries and the Dictionary of Old English (http://www.doe.utoronto.ca), mapping correspondences between these dictionary resources with respect to headword form and coverage. A LangScape lemma was then assigned and mapped to these resources. This forms the core of the markup, allowing us (and also the user) to know exactly what the each lemma represents and thus to apply a standard and consistent LangScape headword-form.

- Separate glossaries have been compiled for the place-names and river-names occurring in the texts, for the Latin headwords, and for personal names, the latter being subdivided into male and female dithematic forms, and monothematic forms, and tagged as to whether or not these names are cited in the Prosopography of Anglo-Saxon England (http://eagle.cch.kcl.ac.uk:8080/pase/index.jsp).

- Each word in the alphabetised list of attested spellings in the boundary texts (13,780 different spellings in total) was mapped as far as practicable to a LangScape headword within the tables. A complex set of programs was then devised to output the texts from the database and convert them into XML files, using the glossaries with their mapped correspondences between textual form and lemma.

- This produced a set of XML files of top texts with every semantically discrete unit tagged as a 'word', and provided with slots for the insertion of lemma and translation. Where correspondences had been established with attested spellings in the Glossary Table, the appropriate lemmas and translations were inserted into the slots. Thus all 8,213 instances of the Old English word 'on', all 5,102 instances of the word 'of' and all 2,376 instances of '7lang' and 'andlang' in our Top Texts, were provisionally lemmatized as 'on', 'of' and 'andlang' respectively, and supplied with the draft translations 'to', 'from' and 'along'. All the texts in the database currently either exist in this provisionally marked-up form, or have proceeded to the second stage of comprehensive in-depth markup (see next section).

- At this point a start was made on systematically preparing each text ('top texts' and 'variant texts') for web publication. In addition, each 'top text' was lemmatized and 'translated', the auto-markup for commonly recurring words releasing more time to be spent on conventional scholarly research with the aid, in

most cases, of the most recently published topographical studies; at this stage, the provisionally marked words could be altered, emended or fine-tuned. To date, 694 'top texts' have been prepared for semi-diplomatic and edited display, and have been lemmatized and translated, and their (609) variant versions prepared for semi-diplomatic and edited display.

- The Glossary Tables sit at the core of the LangScape resource and will allow the user to search the texts via the headwords in these Tables, as well as from string searches on the texts themselves; this is also at an advanced stage of development.

- A mapping facility is currently under development to allow users to plot the estate centres of boundaries containing selected features (topographical or linguistic) onto maps of England, according to further potentially complex criteria selected by the user.

- Links will allow the user to navigate extensively within the database but also to move to external e-resources such as the Electronic Sawyer.

- A 'Learning Old English' Module has been compiled to provide the user with enough basic information on pronunciation and on the structure of boundary clauses to work with the original text, aided by the headword-gloss and 'translation'.

- The boundary clause for Westminster has been used to illustrate how this material can be taken further; it utilises GoogleEarth to lay out the course of the boundary, and walks the user round the byways and waterways of this fascinating London survey. This facility is currently being converted for web-display.

As soon as the complex program which converts the XML files into the different displays has been finalised, the marked-up texts will be given a final check and will be uploaded to the website archive by archive, together with accompanying information over the next few months.

Joy Jenkyns

May 2008