

# eIUS Prototype Use Cases<sup>1</sup>

In these examples, underlined elements are based either on the original interview data or on feedback sent by the informants in response to draft versions of the use cases. Key *activity types* are highlighted in *italics* and **relevant ICTs** in **bold**.

## Use Case 4 - Corpus Linguistics

### Narrative

1. Ella, a corpus linguist based at a large research-led university, with a specialty in modal verbs, receives an email in the early hours from a corpus linguist based in Japan who she has never met before.
2. Karl, who has a particular interest in academic writing wonders whether Ella might be interested in collaborating with him in order to do a comparison of the expression of modality in academic writing compared with British English as a whole. He had just read an article she has just published about modal verbs in school essays. He found it through Google Scholar.
3. The following day, Ella replies to Karl saying she is, in principal, open to such a collaboration, although she makes the point that with limited funds, they are unlikely to be able to meet each other in person in the near future. However, they are not unused to this style of informal collaboration. (Research for both Karl and Ella is an unfunded, out-of-hours activity).
4. They plan to use a **web-based version of the British National Corpus, hosted at Brigham Young University (BYU-BNC)**, which Karl already has some experience with. Ella, on the other hand is more familiar with **BNCWeb**, an alternative web-based interface to the BNC available to members of her university.
5. The next day, Ella visits the **BYU-BNC site**, and phones up Karl so that he can walk her through it. Karl explains the basic search options, and how the search results can be examined in different ways. He also describes how to use some of the more advanced features.

---

<sup>1</sup> Four use cases were produced for the 'Scoping e-infrastructure' interim report, available online at <http://www.eius.ac.uk/scoping/index.xml>. They are reproduced here as separate documents to illustrate how researchers in various academic disciplines use e-infrastructure systems in their work

6. Over the next six months, they divide up the analysis work between them. Ella focuses on analyzing modal verbs (which she has expertise in), whilst Karl works on analyzing other means of expressing modality.
7. For the first month, they analyse and swap data via spreadsheet files sent over email. However, on a recommendation from a colleague of Ella's they trial Google Documents, which incorporates a web-based spreadsheet editing package and allows them to manage versions more efficiently.
8. They decide to present their preliminary findings at a well-known Corpus Linguistics conference. In their talk, some audience members ask questions about whether they have thought about considering whether their findings also hold true in American English, which could shed light on whether academic writing is closer to American or British English.
9. Fortunately, they find out that the same person who developed the **BYU-BNC** is also in the process of developing a comparable corpus of American English. As they are so happy with the way the previous study worked, they contact the developer of the new corpus and he agrees to allow Karl and Ella to beta-test the new system before its official release.
10. After only two months of work using this new corpus, they already have some interesting results. However, the results have to be treated as tentative because the corpus is still only partially constructed. Once it is released, they intend to validate their results with the full corpus and publish their findings in the International Journal of Corpus Linguistics.

## Relevant ICTs

ICT	Comments
Mark Davies/ Brigham Young University Interface to the British National Corpus (BNC) <sup>2</sup>	A publicly accessible web interface to the British National Corpus hosted at Brigham Young University.
BNCWeb <sup>3</sup>	<p>An alternative web interface to the BNC. Equivalent installations of the software (to be released under the GNU GPL<sup>4</sup>) exist at various UK institutions. However, due to BNC licence restrictions, each installation is accessible only to members of the institution hosting the installation. Those with no local installation would need the specialist skills required to install their own copy.</p> <p>The BNCweb interface is considered to be more advanced than the BYU interface. For example, the BNCweb interface allows sorting on terms up to five words away from the search term, can produce ranked lists of collocates, and allows export of data for further analysis offline. However, the BYU interface is publicly available and very fast.</p>
WordNet <sup>5</sup>	The BYU interface to the BNC integrates with WordNet, a semantic analysis of words. For example, it allows searches on all words than mean 'big'.

<sup>2</sup> [corpus.byu.edu/bnc](http://corpus.byu.edu/bnc)

<sup>3</sup> [www.bncweb.info](http://www.bncweb.info)

<sup>4</sup> [www.gnu.org/copyleft/gpl.html](http://www.gnu.org/copyleft/gpl.html)

<sup>5</sup> [wordnet.princeton.edu](http://wordnet.princeton.edu)

## **Commentary**

Much of this use case changed following a review with the original informant who suggested alternative research questions, more appropriate corpora, and more likely (physical) locations for the researchers. What remained the same, however, was the idea of illustrating a geographically distributed collaboration arising between two people who had never met. This feature of the use case was taken from an actual example given by the informant. Also taken directly from the interview data was the description of the research as an unfunded, out-of-hours activity.

There was some debate at the eIUS data session about whether the use of concordance tools represented genuine use of e-Infrastructure. This was because use of BNCweb, considered as e-Infrastructure because it is web-based, did not come out very clearly in the interview data. However, in a follow-up email, the informant confirmed that she 'would not have been able to do her research without BNCweb'. For reasons explained below the use of BNCweb was replaced with use of the similar Brigham Young University interface in the final version of the use case.

The use of Google Documents was included in this example because the informant had indicated their genuine intention to use this tool to improve the management of result data held in spreadsheet files.

## **Comments by Informant**

In this example, the research question, ICTs used, and locations were all changed following review to make the use case more believable from a corpus linguistics point of view.

In the first version the research question, constructed from evidence gathered through desk research rather than from the interview data, concerned the differences between newspaper prose between Manchester and New York in the 1930s. The informant suggested that this was unrealistic because firstly, the BNC was not designed to be representative of UK newspaper journalism but of British English (1970-1995) as a whole and secondly, there is no known corpus of 1930s New York newspaper journalism. As an alternative, the informant suggested that the BNC corpus be used to compare expressions of modality in academic writing compared with British English as a whole. Originally, the BNCweb interface was featured. However, this was replaced with the BYU interface because it is freely available.

The location for Karl was changed from New York because the informant doubted whether a linguist based in New York would be interested in British English and suggested Europe or Japan might be a more likely location, there being several researchers studying British English in these regions. Japan was chosen to highlight the geographically distributed nature of the collaboration.

The use case was also extended by adding the idea of extending the study by carrying out a similar analysis on an American English corpus written by the developer of the BNC interface. However, it is an open question whether including fictional interactions with real people is appropriate, even where such people are referenced only indirectly.

### **Other Editorial Considerations**

<b>Element</b>	<b>Usage</b>
Links to direct quotes?	No
Year?	No
Month?	No
Time of day?	Yes
Location given?	No
Real institutions named?	No
Real journals named?	Yes
Real conferences named?	No