**British National Corpus Baby Word Count**

**CMIST**

**University of Manchester**

**John Mcloughlin**

**September 2016**

1. Extract

Extract from corpus via xsltproc

1. Create Project

In NVIVO, create an empty project, e.g. BNCBaby.nvp.

1. Split Output

NVIVO is unable to load large text files so it is necessary to split the extracted text into 10 files for loading into NVIVO.

1. Load

Load the ten files into NVIVO. Use the Import function. Use External Data > Import Internals menu option in NVIVO.

1. Run Word Frequency Query

Once the word frequency query has been run, the output can be saved to an Excel spreadsheet. NVIVO cannot export more than 65536 rows to an Excel file due to an unknown NVIVO restriction. In this case the output is smaller than the Excel limit, so an output Excel file can be created from the NVIVO word count query.

1. Create Bar Charts

To create bar charts, use grep ( Linux/ Mac ) or FINDSTR ( Windows ) to extract the required word from the CSV file. EG

grep teach BNC\_Baby\_word\_count.csv > teach.csv

Run this command from a Terminal ( Linux/ Mac ) or Command Prompt ( Windows ). This will extract values for all words related to the stem ‘teach’ and create a new CSV file, teach.csv. Load this into Excel and create a bar chart. Save as BNC\_Baby\_teach\_bar.xsls