The Talk of the Toon

http://research.ncl.ac.uk/decte/toon

Newcastle University

http://www.ncl.ac.uk

The Diachronic Electronic Corpus of Tyneside English

http://research.ncl.ac.uk/decte

DECTE is funded by:

Arts & Humanities Research Council

http://www.ahrc.ac.uk

Exploring *The Talk of the Toon* / DECTE interviews

Exercises

A. Using The Talk of the Toon to explore the DECTE interviews

http://research.ncl.ac.uk/decte/toon

The Talk of the Toon site presents the DECTE interviews in the form of a multimedia online archive, by linking the recordings and transcripts, and combining them with thematically related pictures and video clips.

The archive's THEMES page is accessed by selecting 'Themes' in the navigation menu under the banner at the top of any page, or clicking 'More' in the THEMES panel on the homepage. The menu on the left of the THEMES page lists the seventeen available themes. Clicking one of these brings up the results summary for that theme. This shows the number of associated pictures and video clips, and the number of interviews containing relevant key words.

- Click 'Expand photo results' or 'Expand video results' to view the pictures and clips respectively.
- The interview list can be filtered (by time period, and by informant age and gender) using the selection menu and 'Filter results' button under the results summary window.
- Click 'Text & Audio of full interview' to go to any interview's Text/Audio page (see below). The theme's key words will be highlighted, and the transcript will show the first example in the text.

As well as accessing an interview's Text/Audio page via the THEME results panel, you can also view an index of all the interviews in the archive by clicking 'View list of all interviews' at the top of the THEMES menu, or by choosing 'Interview Index' from the navigation menu under the banner at the top of a page. The INTERVIEW INDEX page lists files in sets according to interview date, and gives details of the gender and age of the informant(s). Click on an interview link to go to the relevant Text/Audio interface page. The interface allows you to:

- Click any section of the transcript to begin playback of the audio at that point in the interview.
- Select a point in the audio timeline to jump to that part in both the recording and transcription.
- Highlight key words in the interview belonging to one of the seventeen themes listed in the THEMES panel next to the transcription window.

A.1. Themes

a. On *The Talk of the Toon* THEMES page, Select **SPORT** from the menu on the left-hand side of the screen.

b. The Photo, Video and Text/Audio results for **SPORT** appear in the results panel.

c. The Text/Audio window lists all interviews in which key words relating to the **SPORT** theme occur, indicating the gender and age of the informant(s), and time period of the interview. Currently, the archive contains 37 DECTE interviews from the 1960s-70s, 18 from the 1990s, and 43 from the 2000s. There are 86 female informants (20 in the 1960s-70s set, 18 from the 1990s, and 48 from the 2000s) and 72 male informants (17 in the 1960-70s set, 17 from the 1990s, and 38 from the 2000s).

d. Below the results panel is a filter menu. Click on the box next to one or more of the options (a ✓ indicates selection). Then click 'Filter Results'. The Text/Audio results panel now lists only those interviews that meet the chosen criteria, e.g. only interviews from the 1960s-1970s, and/or only interviews with people of a particular age group, etc.

e. Is **SPORT** a more popular theme in the interviews from the 1960s-70s, the 1990s, or the 2000s? (You need to think about the total number of interviews from each time period.)

f. In relative terms, is it a more popular theme in conversations involving speakers up to and including 30 years of age (88 speakers in total) or the over 50s (38 people)?

Repeat the process with some of the other themes, e.g. **ENTERTAINMENT & CULTURE**, etc.

A.2. Interview Interface

Speakers represented in DECTE naturally have different opinions and experiences. One way of comparing them is to use *The Talk of the Toon* interface to highlight a THEME in interviews that involve different kinds of informants, or come from different periods. For example: From the INTERVIEW INDEX page, select the final interview from the 1960s-70s **(TLSG37)**.

a. Click on **HOLIDAYS** in the THEMES panel on the right-hand side of the page. Key words related to this theme are highlighted in the interview, and the text jumps to the first use of one of these key words.

b. Click on the text where the first key word appears ('*what about holidays*'). The section of text containing the key word is highlighted and the audio begins to play, starting at the section you have selected.

c. You can use the scroll bar on the right-hand side of the interview text to find other highlighted key words in the transcription.

d. Now try the **ENTERTAINMENT & CULTURE** theme.

e. Compare the references to 'HOLIDAYS' and 'ENTERTAINMENT & CULTURE' in **TLSG37** with those in e.g. **TLSG11** (also 1960s-70s) and **PVC13** (1990s).

Try the same procedure with other themes and interviews. You can go back to one of the THEME results pages you were looking at and choose one of the interviews listed there.

You can also compare what informants have to say on a given topic with the views expressed in the news reports found among the video results for the various THEMES.

As well as selecting a theme or clicking on a section of text, you can of course simply click on the play button (▶) on the audio player, to listen to an interview from the beginning.

B. Analysing the DECTE text files

As well as forming the core of *The Talk of the Toon* website, the DECTE interviews are available for download as text files (see the 'Corpus Files' section of the DECTE website for details). These files can of course be read and searched individually, using any standard word processor or text editor program. However, they can also be searched in batches or as a complete set using dedicated text analysis software.

The exercises below involve searching the files with the freeware program AntConc (http://www.antlab.sci.waseda.ac.jp/antconc_index.html). A guide to using AntConc with the DECTE text files is available in the 'Schools' section of *The Talk of the Toon* website.

B.1. Language and Gender

Andrew Moore outlines some of the claims that scholars have made about the distinctive characteristics of male and female speech (www.teachit.co.uk/armoore/lang/gender.htm).

Select some DECTE interviews that involve mixed and same gender pairs (see the 'About' page on *The Talk of the Toon* site for details of interview informants).

Using AntConc, search for appropriate words and phrases to investigate some of the claims about male and female speech. For example:

- Lakoff 1975: women tend to use more hedges (*sort of*, *kind of*, *seems like*) and apologise more (*sorry*).
- Dominance Theory: men interrupt more than women. To investigate this, you can search the text files for the interruption tag: *<interruption>*. Consider also the distinction that Tannen makes between 'competitive overlap' (i.e. interruption) and 'cooperative overlap'.
- Coates 1998: 'Lexical items such as *perhaps, I think, sort of, probably* ... are used in English to express epistemic modality ... women use them to mitigate the force of an utterance in order to respect addressees' face needs.'

With any investigation of this kind using text analysis software, you will need to look carefully at the results you get, to determine which are relevant to the focus of your study.

For example, searching for 'kind of' as part of a study of hedges will also return results where it has a different function, e.g. in questions like 'so what kind of car would you get' (decten1pvc09, line 415).

B.2. Language Features: Intensifiers and Quotatives

a. Barnfield & Buchstaller (2010) traces changes in the frequency and usage of various words that are used as intensifiers in the DECTE interviews from the 1960s-70s, 1990s and 2000s (*really bad*, *very good, dead posh*, etc).

Barnfield, Kate & Isabelle Buchstaller. 2010. 'Intensifiers on Tyneside: Longitudinal developments and new trends'. *English World-Wide* 31 (3): 252-287. For a summary, see: http://linguistics-research-digest.blogspot.co.uk/2011/11/reallyverydeadso-interesting-how-to.html

- Use AntConc to search for different intensifier forms in the DECTE files, e.g. **very, really, dead, so, pure**. Which are most common in each of the three periods represented in DECTE?
- Are there any problems involved in searching for contexts in which these and other terms are used as intensifiers?

b. Recent research on DECTE has shown how the quotatives BE LIKE (*she was like "why?"*) and GO (*I went "oh typical"*) emerged in the 1990s interviews, and became increasingly frequent in the 2000s.

Buchstaller, Isabelle. 2011. 'Quotations across the generations: A multivariate analysis of speech and thought introducers across 5 decades of Tyneside speech'. *Corpus Linguistics and Linguistic Theory* 7 (1): 59-92. For a summary, see: http://linguistics-research-digest.blogspot.co.uk/2012/03/quoting-across-generations.html

- Use AntConc to search for different quotative forms.
  For example: **I'm like** / **I was like** / ***he's like** / ***he was like** (**he* will find both *he* and *she*) and **I went** / **I goes** / ***he went** / ***he goes**, etc.
- Can you find any patterns in the way that the quotatives are used by speakers of different genders or ages, or in terms of the kinds of quote they introduce?

B.3. Miscellaneous

a. In the DECTE interviews from the 1960s-1970s (the *Tyneside Linguistic Survey* recordings), the interviewer often asks people about their political views.

- Use AntConc to search the DECTE interviews for **vot*** (which will find *vote*, *votes*, *voted*, *voting*, *voter*, *voters*) and **politic*** (*politics*, *political*, *politically*, *politician*, *politicians*).
- Is politics also a frequent topic of conversation in the 1990s and 2000s? How do the conversations and/or the use of the key terms differ across time?
- What other key terms could you search for to extend your study of this semantic field?

b. David Bellos notes that 'when we say 'It was literally raining cats and dogs last night', we mean the adverb 'literally' in a figurative sense. Studies of large corpora of recorded speech have shown that the majority of the uses of 'literal' and 'literally' in English are figurative.'

Bellos, D. 2011. *Is That a Fish in Your Ear? Translation and the Meaning of Everything*. Penguin: London. Page 110.

- Use AntConc to search for **literal*** (*literal*, *literally*). Do the occurrences of these words in the DECTE interviews support the point that they are mostly used in a figurative way in English?

C. Some other online resources

C.1. ENROLLER

http://www.gla.ac.uk/enroller

ENROLLER is an integrated online repository, allowing cross-searching of multiple resources. Currently, two of the component parts of DECTE are part of this repository: the interviews from the 1960s-1970s (TLS) and from the 1990s (PVC), which were digitized and combined between 2000-2005 to form the ***Newcastle Electronic Corpus of Tyneside English*** **(NECTE)**.

- From the ENROLLER homepage, select '**Portal**' in the menu on the left-hand side of the screen. You are presented with a search box.
- Enter a search term, e.g. **canny**, and press ENTER or click 'New Search'.
- The results list indicates which of ENROLLER's data sets contain examples of the search term.
- You can examine results from a particular data set by clicking on either 'Show details' (to expand the relevant section) or 'Show results in a window' (to see the relevant entries in a separate pop-up window).

For investigations that focus on the use and meaning of dialect words, one useful approach is to search both the NECTE data set and the *Historical Thesaurus of English* (HTE).

- Try searching for the following traditional North East dialect words: **bairn** (child), **bonny** (pretty; fine), **boody** (piece of broken chinaware), **clarts** (mud), **canny** (fine, good, admirable; quite, very), and **gan** (to go).
- Does the HTE contain entries for the all of these words?
- Where the HTE does contain entries for these words, do they (always) include the meanings that are relevant to the contexts in the NECTE interview results list? (NOTE: You can get a pop-up window with more details on the various words that the HTE records for a particular meaning by clicking on the blue code number that appears next to that meaning in the Thesaurus results list.)
- It is also worth thinking in more detail about the contexts in which some of these words appear in the NECTE results. Are they always part of the normal flow of conversation? In the 1960s-1970s (TLS) interviews in particular, who introduces the words: the informant or the interviewer?
- Try searching for some other words, and/or cross-searching NECTE and some of the other ENROLLER resources, e.g. SCOTS (The Scottish Corpus of Texts and Speech).

C.2. SCOTS (The Scottish Corpus of Texts and Speech)

http://www.scottishcorpus.ac.uk

a. As indicated on the website's 'Details' page, SCOTS is a corpus of over 4 million words, consisting of material from over 1100 written and spoken texts, covering the period from 1945 to the present. In addition to Scots and Scottish English, there are a small number of texts in Scottish Gaelic.

There is a 'Quick Search' box on the homepage.

There are a number of words that the *Oxford English Dictionary* identifies as belonging to both Scottish and Northern English varieties.

Try searching SCOTS for the following words. Judging by the contexts in which they are used, do they have the same meaning(s) in the Scottish texts as they do in the dialect of the North East of England?

- **bonny** (pretty; fine), **clarty** (muddy), **canny** (fine, good; quite, very), **neep** (turnip), **polis** (police)

b. The SCOTS Collocate Cloud:
http://www.scottishcorpus.ac.uk/corpus/search/collocatecloudstart.php

The SCOTS Collocate Cloud presents a visual representation of up to 100 of the words that are most commonly found near the chosen key word – i.e. the most frequent collocations.

- Try generating clouds for some of the words mentioned above (*bairn*, *canny*, etc), or pick your own.
- The larger a word in the cloud, the more frequently it occurs in the vicinity of the key word. Brighter words have a stronger association with the key word (i.e. more likely to occur *usually* or *only* in association with the key word ... like "Dec" tends to occur only in the same context as "Ant").
- Notice that not all of the listed words appear to be relevant. Some evidently relate to additional information contained in the texts (e.g. "[laughter]" or number codes).

C.3. The BNC (British National Corpus)
http://www.natcorp.ox.ac.uk

From the BNC homepage: 'The British National Corpus (BNC) is a 100 million word collection of samples of written and spoken language from a wide range of sources, designed to represent a wide cross-section of current British English, both spoken and written.'

- Try doing BNC searches for some of the words you've already searched for using ENROLLER and SCOTS.

The BNC Collocate Cloud (provided by SCOTS): http://www.scottishcorpus.ac.uk/corpus/bnc

The BNC Collocate Cloud does the same thing as the SCOTS Collocate Cloud (see C.2.b), but with the BNC files.

- Compare the results you got from the SCOTS Collocate Cloud with BNC clouds for the same words.