A corpus approach to cultural keywords: a critical corpus-based analysis of ideology in the Blair years (1998-2007) through print news reporting

Lesley Jeffries
l.jeffries@hud.ac.uk
The University of Huddersfield, UK

Brian David Walker
b.d.walker1@lancaster.ac.uk
The University of Huddersfield, UK

This paper will report on a corpus-based study of the cultural keywords (in the Raymond Williams' sense) via the analysis of key-words (in the corpus/statistical sense) of newspaper reporting in the years since Labour came to power. The project demonstrates that certain lexemes (or lexical strings) gain currency in relatively short historical periods and may take on political importance.

This project assesses the ideological landscape during the years of the New Labour project by extracting the cultural keywords of the time, and demonstrating their evolving meanings in the commentary provided by the print media.

The project takes inspiration from Raymond Williams' book ([1975] 1983) *Keywords* which attempted to sum up the ideology of the post-war years. Williams chose a set of words which he thought had taken on particular meanings in that period, and wrote an informed but ultimately anecdotal commentary on each one. Like Williams, we begin with a hypothesis that some words (such as, for example, *radicalisation*, *choice*) have both increased in usage and polarised in their meaning since 1998.

Unlike Williams, this project pursues a rigorous approach to the discovery of *which* words characterise the period under investigation, using two corpora of newspaper data and computer tools. This enables us to make a comprehensive investigation and an objective assessment, including use and meaning, of the cultural keywords of the Blair years

The project is primarily corpus-based, but with a strong qualitative focus, using an approach to studying textually-constructed meanings of words and other linguistic items which recognises both their place in a relatively stable system of language, and their capacity to take on additional meaning in specific contexts of time and place.

1. Background

Our project links the corpus linguistic notion of key-words to earlier work into the 'emergent meaning' of individual lexical items (see Jeffries 2003, 2006 and 2007).

Jeffries (2003) investigated the meaning of water, in the context of the Yorkshire water crisis of 1995. Jeffries (2006) investigated the speech act of apology, in particular news commentators' view of Blair's putative apology for the Iraq war. Jeffries (2007) was a much more extensive consideration of the way in which the female body was constructed by women's magazines in 2004. This larger study developed a system of describing textual meaning which draws on Hallidayan approaches to systems of linguistic form and meaning applying his combined semantic and syntactic view of textual meaning to other functions such as the construction of opposites.

The current project was designed in the spirit of Critical Discourse Analysis, in particular the work of Fairclough whose work on ideology in language, and specifically the language of New Labour (Fairclough 2000) influences the approach taken here. However, the methods used in this project are closer to corpus stylistics in that they are text-analytic and at least in some of the stages, computer-assisted and corpus-driven. Work already carried out in this area (see for example McIntyre and Walker 2010) showed that corpus approaches and tools, in particular Wmatrix, can successfully be applied to textual analysis. Baker and McEnery (2005) and Baker and Gabrielatos (2008) are also influential on this project because this work has paralleled Jeffries' work in looking at sets of texts from a particular time period to demonstrate political ideologies in news texts.

The project also reflects renewed interest in cultural keywords in the Williams sense, with a recent special issue of Critical Quarterly (2007) devoted to the subject, and Durant's (2006) related article which suggests that "[...] the development of electronic search capabilities applied to large corpora of language use [...] encourages renewed attention to cultural keywords." (Durant 2006). This project effectively takes up that suggestion.

2. Research questions

1. What are the key-words for the years 1998–2007, as evidenced in the British press and can they be identified as cultural keywords?

2. Have they developed meanings specific to this period and have these meanings evolved within the period?

3. Methodology

The project focuses on news texts from 1998 to the end of 2007. A corpus of comparable data from three national daily newspapers (The Guardian, The Independent, and The Times) was assembled from a large, on-line newspaper database. This database represents a very rich and potentially overwhelming amount of data (100s of millions of words). However, our project had very limited timescales and we found it necessary to carefully control the amount of data that we collected. This was because: (i) downloading selected articles from the database is largely a manual and fairly time consuming process; and (ii) in its raw form each downloaded article contained structured extra-textual details (headers containing titles, dates and so forth), random intra-textual information (such as journalist's email addresses) and corruptions. This extra text and corrupted data had to be removed from and amended in each of the downloaded files: a laborious process which consumed a lot of project time. We also found that the corpus tools we used for data manipulation struggled when presented with files of more than one million words. Consequently, we took a structured sampling approach, choosing a week from the politically 'busy' month of September (party conferences), and collected selected news-related items from these weeks. The resulting corpus was approximately 2.3 million

words, which we anticipated would be sufficient to answer our research questions. A comparison corpus was built along similar lines using newspaper data from the five year period prior to 1997 (the Major years).

The corpus was automatically analysed, in the first instance, using Wmatrix (Rayson 2008), which is a relatively new corpus tool that can calculate keyness (using Log-likelihood) at the word level (key-words), at the grammatical level (key-POS), and the semantic level (key-concepts). The present study uses just the key-word output.

To address the qualitative aspect of the research questions, this investigation included the following considerations:

- Do the collocations of the key-word demonstrate particular nuances of meaning?

- How does the semantico-syntactic behaviour of the key-word demonstrate meaning specific to the context?

- Does the key-word enter into any unconventional lexical relations (e.g. of opposition)?

- Is the key-word associated with any modal or negated text worlds?

4. Results

The key-words, generated from the comparison of our corpora, that we consider to be the important cultural keywords from the Blair years are as follows:

No.	Keyword	Associated key-words
1	Terror	Terrorism, terrorist(s), attacks, atrocities, threat
2	Global	Globalisation, world, international
3	Spin	spun
4	Reform	progressive, radical, modernise(d) / er(s) / ation
5	Choice	
6	Respect	

Items in the 'keyword' column are the main items used in our investigation and the terms that we consider to be culturally significant. The items in the third column are key-words resulting from our corpus comparison which are related to individual (cultural) keywords and which, we hypothesise, form a network of

meaning. These are still to be fully investigated and we do not report on them in this paper.

For each keyword we provide a more detailed quantitative commentary using concordance and collocation data. Our major findings, though rigorous and replicable, are qualitative, and provide the basis of both detailed linguistic commentaries on each key-word and could also provide the foundation for more general popular essays not dissimilar to those provided by Williams, but with more clarity about their provenance. There will not be time to discuss all our findings, but our paper will report on some of the quantitative data and focus qualitatively on 'spin'.

References

Adamson, S., Durant, A (2007). *Critical Quarterly*. **49,1**.

Baker, P., McEnery, A. (2005). 'A corpus-based approach to discourses of refugees and asylum seekers in UN and newspaper texts'. *Language and Politics*. **4:2**: 197-226(30).

Baker, P., Gabrielatos, C. (2008). 'Fleeing, sneaking, flooding: a corpus analysis of discursive constructions of refugees and asylum seekers in the UK Press 1996-2005'. *Journal of English Linguistics*. Forthcoming.

Durant, A. (2006). 'Raymond Williams's Keywords: Investigating Meanings "offered, felt for, tested, confirmed, asserted, qualified, changed"'. *Critical Quarterly*. **48:1**: 1–26.

Fairclough, N. (2000). *New Labour, New Language*. London: Routledge.

Jeffries, Lesley (2003). 'Not a drop to drink: Emerging meanings in local newspaper reporting of the 1995 water crisis in Yorkshire'. *Text - Interdisciplinary Journal for the Study of Discourse*. **23 (4)**: 513-538.

Jeffries, Lesley (2006). 'Journalistic Constructions of Blair's 'Apology' for the Intelligence Leading to the Iraq War'. *Language in the Media: Representations, Identities, Ideologies. Advances in Sociolinguistics*. London: Continuum.

Jeffries, Lesley (2007). *Textual Construction of the Female Body A Critical Discourse Approach*. Basingstoke.

McIntyre, D., Walker, B. (2010). "How can corpora be used to explore the language of poetry and drama?". *The Routledge Handbook of Corpus Linguistics*. McCarthy, M., O'Keefe, A. (eds.). Abingdon: Routledge.

Rayson, P. (2008). *Wmatrix: a web-based corpus processing environment,* Computing Department, Lancaster University `http://www.comp.lancs.ac.uk/ucrel/wmatrix/`.

Williams, R. (1983). *Keywords (2nd Ed.)*. London: Fontana.