

# A Word on the Wellington Corpora<sup>1</sup>

Bernadette Vine

Victoria University of Wellington

## 1. Introduction

Three one million word corpora for exploring and understanding the nature of New Zealand English (NZE) are now available at the School of Linguistics and Applied Language Studies, Victoria University of Wellington. These three resources are The Wellington Corpus of Written New Zealand English (WWC), The Wellington Corpus of Spoken New Zealand English (WSC) and the New Zealand Component of the International Corpus of English (ICE-NZ). Preliminary research on a wide range of topics has already been undertaken on data drawn from these three corpora, although there are many other areas still to be explored. This paper will provide a brief summary of the corpora and of some of the initial findings in terms of NZE vocabulary.<sup>2</sup>

## 2. The Wellington Corpora

### 2.1 The Wellington Corpus of Written New Zealand English (WWC)

The first of the corpora to be completed at Victoria University was the WWC. The WWC was collected from writings published in the years 1986 to 1990 and consists of 2,000 word excerpts on a variety of topics. Text categories include press material, religious texts, skills, trades and hobbies, popular lore, biography, scholarly writing and fiction. These categories were designed to parallel the Brown Corpus of written American English (1961), the Lancaster-Oslo-Bergen corpus (LOB) of written British English (1961) and the Macquarie Corpus of written Australian English (1986) (see Bauer 1993 for further information).

### 2.2 The Wellington Corpus of Spoken New Zealand English (WSC)

The WSC was completed in mid-1998. It consists of 2,000 word extracts (where possible) and comprises different proportions of formal, semi-formal and informal speech. Both monologue and dialogue categories are included and there is broadcast as well as private material collected in a range of settings. Most of the material included is informal dialogue with 75% of the one million words falling into this category (see Holmes, Vine and Johnson 1998 for further details).

---

<sup>1</sup> We are very grateful to the many people who contributed their voices, time and expertise to the corpora projects over the last ten years.

<sup>2</sup> If anyone would like additional information about the information in this paper or details on the availability of the corpora please contact the Manager at [Corpus-Manager@vuw.ac.nz](mailto:Corpus-Manager@vuw.ac.nz) or Corpus Manager, School of Linguistics and Applied Language Studies, Victoria University of Wellington, PO Box 600, Wellington.

## 2.3 The New Zealand Component of the International Corpus of English (ICE-NZ)

The latest corpus to be completed at Victoria University is ICE-NZ. ICE-NZ consists of 600,000 words of spoken material and 400,000 words of written material. It is only the second component of the International Corpus of English to be made available. Within the next few years, another 16 ICE components collected in different countries around the world will be completed. These all have the same basic categories.

## 3. Some Initial Research Findings on NZE Vocabulary

### 3.1 Maori in NZE

Ongoing work by Graeme Kennedy on the WSC and the WWC involves an investigation of the use of Maori in NZE. He has found that about six in every 1000 words is of Maori origin. Most of these words are place names and personal names rather than words referring to culturally specific objects or concepts. Another interesting finding of his research so far has been the number of Maori words in the WSC which do not appear in the Dictionary of New Zealand English (Orsman 1997) and vice versa.

### 3.2 Sexist Language

Another vocabulary area explored has been the use of sexist language in NZE (Holmes 1993a, 1993b, 1993c, 1999). In these investigations Janet Holmes has analysed a number of different terms, most recently the use of *woman/women* and *lady/ladies* in the ICE-NZ and ICE-GB corpora<sup>3</sup> (Holmes 1999). Holmes set out to explore Lakoff's (1975) claim that *woman* will be replaced by *lady* because of the sexual connotations of the term *woman*. She found that New Zealanders are using *woman* more and that *lady* tends to occur in set expressions such as *ladies and gentleman*.

### 3.3 Talking about 'time' in Spoken Discourse

The ways people talk about time was recently investigated by Uta Lenk in order to explore the usefulness of corpora as a tool for looking at set expressions and phrases (Lenk 1999). Lenk drew her data from several international spoken corpora, including the WSC. The corpora proved a valuable resource for exploring expressions involving a reasonably high frequency word such as *time*. She found that people *spend time* and *waste time* frequently and are often *given a hard time*, but very seldom *give other people a hard time* (Lenk 1999).

## 4. Conclusion: The Corpora as Research Tools

The brief outline above of some research on the words New Zealanders use demonstrates the range of uses for corpora such as the Wellington corpora,

---

<sup>3</sup> ICE-GB is the British Component of the International Corpus of English. This was the first component of the International Corpus of English to be completed.

even within just one area of investigation. Research has also been conducted in many other areas, for example, Jen Hay's (1995) study of gender and humour, Janet Holmes and Maria Stubbe's (1995) investigation of pragmatic devices, Robert Sigley's (1997) exploration of relative clauses and Wen Yang's (1996, 1997) discourse analysis of direct and indirect speech.

The two Wellington Corpora and ICE-NZ provide an invaluable resource for linguistic research. The New Zealand focus enables descriptions of spoken and written New Zealand English to be carried out, while the large proportion of informal data in the WSC provides material for investigations of everyday speech. ICE-NZ is an excellent resource for comparative studies of varieties of English, especially as more components of the International Corpus of English are completed and become available.

## 5. References

- Bauer, Laurie 1993. *Manual of Information to accompany the Wellington Corpus of Written New Zealand English*. Wellington: Department of Linguistics, Victoria University of Wellington.
- Hay, Jen 1995. *Gender and Humour: Beyond a Joke*. Unpublished MA (Linguistics) thesis. Wellington: Victoria University of Wellington.
- Holmes, Janet 1993a. Chairpersons and goddesses: non-sexist usages in New Zealand English, *Te Reo* 36: 99-113.
- Holmes, Janet 1993b. He-man beings, poetesses, and tramps: sexist language in New Zealand. In Laurie Bauer and Christine Franzen (eds), *Of Pavlova, Poetry and Paradigms: Essays in Honour of Harry Orsman*. Wellington: Victoria University Press.
- Holmes, Janet 1993c. Sex-marking suffixes in New Zealand written English. *American Speech* 68, 4: 357-370.
- Holmes, Janet 1999. 'Ladies' and 'Gentlemen': Corpus Analysis and Linguistic Sexism. Paper presented at the 20th Annual Meeting of the International Computer Archive of Modern/Medieval English (ICAME 1999), Freiburg, Germany, May 1999.
- Holmes, Janet and Maria Stubbe 1995. You know, eh and other 'exasperating expressions': an analysis of social and stylistic variation in the use of pragmatic devices in a sample of New Zealand English. *Language and Communication* 15: 63-88.
- Holmes, Janet, Bernadette Vine and Gary Johnson 1998. *Guide to the Wellington Corpus of Spoken New Zealand English*. Wellington: School of Linguistics and Applied Language Studies, Victoria University of Wellington.
- Lakoff, Robin 1975. *Language and Woman's Place*. New York: Harper & Row.
- Lenk, Uta 1999. Stabilized Expressions in Spoken Discourse: Worth our Time? Paper presented at the 20th Annual Meeting of the International Computer Archive of Modern/Medieval English (ICAME 1999) Freiburg, Germany, May 1999.
- Orsman, H W 1997. *The Dictionary of New Zealand English*. Auckland, NZ: Oxford University Press.
- Sigley, Robert 1997. *Choosing your Relatives: Relative Clauses in New Zealand English*. Unpublished PhD Thesis. Wellington: Victoria University of Wellington.
- Yang, Wen 1996 *Discourse Analysis of Direct and Indirect Speech in Spoken New Zealand English*. Unpublished MA (Applied Linguistics) thesis. Wellington: Victoria University of Wellington.
- Yang, Wen 1997. Discourse Analysis of direct and indirect speech in spoken New Zealand English, *New Zealand Studies in Applied Linguistics* 3, 62-78.