The Maori lexical presence in New Zealand English: Constructing a corpus for diachronic change

John Macalister[1]

Abstract

This paper reports on the construction of a corpus designed to measure changes in the presence of Maori words in New Zealand English over the 150-year period from 1850 to 2000. It begins with a brief introduction to the variety, describes issues identified prior to the commencement of the corpus's construction, and discusses ways in which those issues were addressed.

1. Introduction

1.1 Origins and development of New Zealand English

New Zealand English is a relatively new variety of English. Until 1769, when Captain James Cook made landfall, te reo Maori, the Maori language, was the sole language of New Zealand. Cook and his crew did not settle, but by the early nineteenth century small numbers of English speakers, including missionaries, traders and whalers, had established themselves in New Zealand. When in 1840 the Treaty of Waitangi was signed between a representative of the British Crown and numbers of Maori chiefs, around two thousand non-Maori were residing in the country. This treaty, regarded as New Zealand's founding document, signalled the beginning of systematic colonisation, and within twenty years the English-speaking population outnumbered the indigenous Maori-speakers. By the end of the century commentators were remarking on both phonological and lexical features of the newly emerging variety of English (McBurney, 1887; Morris, 1898). Over the next hundred years the distinctive features of New Zealand English would become increasingly clearly defined.

The origins of the variety lie both in its British, and particularly English, roots (see Gordon *et al.*, 2004) and in the contact between the Maori and English languages. The latter is particularly evident in the lexicon. Deverson (1991: 18) identified lexical borrowing from te reo Maori as being 'in reality the most unmistakably New Zealand part of New

---

[1] *Correspondence to*: John Macalister, *e-mail*: john.macalister@vuw.ac.nz
address: Von Zedlitz Building, Kelburn Parade, Kelburn Campus, Victoria University of Wellington, Wellington 6001, New Zealand

Zealand English'. He also noted that 'Maori is making its presence in English more strongly felt than ever before' (1991: 19) as a result of social change, although this claim was based on observation and impression of linguistic performance rather than empirical data. The compilation of two corpora of New Zealand English, one written, the other spoken (Bauer, 1993; Holmes *et al*., 1998), allowed quantitative analysis of the Maori presence in contemporary New Zealand English (notably Kennedy and Yamazaki, 1999; Kennedy, 2001), but diachronic change could not be measured. It was to measure diachronic change in the presence of Maori words in New Zealand English and to test claims about an increasing Maori presence in the lexicon that this corpus was created.

2. Issues in construction of the corpus

This corpus was intended to provide information about changes in the presence of words of Maori origin in standard written New Zealand English over a 150-year period. The key issues concerning validity and reliability were ensuring that the corpus would be representative of the variety under investigation and constructing the corpus in such a way as to ensure that diachronic comparisons could be made with confidence. These issues were addressed through considering:

- the management of the time frame
- the selection of sources
- the comparability of the data
- the storage and analysis of data

2.1 Management of the timeframe

As this project surveyed lexical borrowing over a period of 150 years, and sought to establish trends within that period, an early decision about a tool to manage the timeframe was necessary. It would not do, for example, simply to select years that were likely to have significant Maori content. Instead, the method employed was to sample the extent of borrowing within the selected data sources every thirtieth year, beginning in 1850. These years are referred to as "indicator years". An indicator year allows the researcher to take a synchronic snapshot of the language in use at a given time, and, by comparing a series of these snapshots, to establish diachronic trends. This assumes, however, the comparability of the data sources in each of the indicator years and is discussed below. Thirty years was chosen as the gap between each indicator year as thirty years is generally accepted as being the measure of a generation (OED). Thus, by employing six indicator years (1850, 1880, 1910, 1940, 1970, 2000), it

would be theoretically possible to measure the development of New Zealand English over six generations.

2.2 Selection of sources

Having established a means for managing the time-span, consideration then needed to be given to the selection of sources. To ensure the reliability of the findings, it would be necessary to compare like with like. To achieve this, source data that had been available throughout the period 1850 to 2000 needed to be identified. The most likely candidate for such data was the print media, which would also be a suitable source as newspapers speak to their readers in their readers' language, and so present an adequate representation of the language in use at a given time.

Newspapers

Printing presses accompanied many of the early European colonists to New Zealand, and English language newspapers soon appeared in the centres of settlement, as the following extracts from Scholefield (1958: 1-2), make clear.

> 'Samuel Revans had no sooner stepped ashore at Port Nicholson [than] he set up his press on the Petone beach, …'

> 'Canterbury, the last of the six original settlements, had to wait for its paper only long enough for the plant to be unloaded at Lyttelton.'

> 'New Plymouth was an exception to the general rule. Though it was settled in 1841 this tortured little province was so retarded by native land troubles that the people were without a newspaper till 1852.'

Issues of representativeness needed to be considered before selecting the actual newspapers to be used, however. These issues can be summarised as:

- the balance between metropolitan and provincial newspapers
- the balance between North and South Island newspapers
- the balance between Maori and English place names; if all the newspapers came from cities/provinces with Maori names, this would likely lead to an inflation of Maori toponyms, just as selection from cities/provinces with English names would probably cause a serious under-representation
- the consideration of historical continuity of publication; as Bell (1991: 18) points out, 'Gathering historical media language data presents particular problems for continuity'. The ideal is to use newspapers that have published continuously throughout the

period under investigation, but even such newspapers can merge, change name, change ownership, and certainly change editors. However, while such changes may impact on a newspaper's style, the very fact of commercial survival suggests that they are in harmony with the reading public's expectations and practice.

The first two issues relate to demographic concerns. There needed to be some assurance that the source data was representative of the entire English-language-using population of New Zealand in any given indicator year. If, for example, a South Island newspaper was presented as the sole source for media language in 2000, when two-thirds of the country's population resides in the North Island, with more than a million people in the Auckland region alone, the data could lack face validity.

In the end, four daily newspapers with a continuous history of publication were chosen for the indicator years 1880, 1910, 1940, 1970 and 2000:

- the *New Zealand Herald* (North Island metropolitan, English place name)
- *The Dominion* (North Island metropolitan, English place name)[2]
- the *Otago Daily Times* (South Island metropolitan, Maori place name)
- the *Wanganui Chronicle* (North Island provincial, Maori place name)

For the indicator year 1850, partly because other metropolitan newspapers still in existence had not yet begun publication at that date, the entire run of the *Wellington Independent* for that year was analysed (see footnote 2, and further discussion below). Also, the distinction between metropolitan and provincial newspapers had no meaning in 1850, as no European settlement had a population greater than a few thousand at that time. Auckland, the most populous centre of settlement, had a European population of 8,301; the Otago settlement numbered 1,482 (Grey, 1994: 186). The entire European population of New Zealand in 1851 was only 26,581 (Grey, 1994: 164).

Further decisions needed to be made following the selection of newspapers. One such decision concerned the content to be included. Newspapers consist of a variety of material, the two dominant genres being

---

[2] The history of newspapers in Wellington is slightly more complicated than that of the other main centres. The *Wellington Independent* ran from 1845–1874, when it was purchased by the *New Zealand Times*, which continued the *Independent*'s serial numbering. The *New Zealand Times* co-existed with *The Dominion*, which was established in 1907, for twenty years until being incorporated into the latter in 1927. Since this research was undertaken *The Dominion* has changed yet again, having merged in 2002 with its evening rival to become *The Dominion Post*. Thus, this corpus drew on the *Wellington Independent* in 1850, the *New Zealand Times* in 1880, and *The Dominion* for the twentieth century indicator years.

news and advertising (for a full discussion see Bell, 1991: 12–15). In the construction of this corpus the advertising genre and the news sub-category of service information (for example, sports results, television programmes) were excluded in their entirety. The reasons for this decision were that in advertising, language is often used either creatively and unusually or formulaically, and that service information tends to consist of lists. One additional qualification was applied to inclusion in the corpus. The research was focussing on written New Zealand English, and so a working definition of New Zealand English, as it appears in the print media, needed to be applied. This is a problem which has vexed other corpus compilers (Holmes, 1995: 6-7; for example), although for this corpus the origins of language-users were not considered to be a source of concern, as the research focus was lexical. Topic and audience were presumed to determine the vocabulary used, even in 1850 when the news was, inevitably, written by recent migrants whose English variety would have been British. As a result, on this occasion New Zealand English was defined as any piece of news or opinion that was written in New Zealand about New Zealand. 'About New Zealand' was broadly defined to include occurrences within the country, even, for example, interviews by New Zealand journalists with international visitors. Nevertheless, this automatically excluded almost all international news, and all feature articles sourced from overseas, where distinctly New Zealand words were not expected to be found.

Having decided what content areas in the newspapers were to be included or excluded, the next major issue was how much to be included. One of the potential dangers in using media language is being overwhelmed by the sheer quantity of available data. Given that a single issue of a major New Zealand metropolitan newspaper may contain anything between one and two hundred thousand words, an entire year of publication is going to make tens of millions of words available for analysis. However, a representative sample of the entire year can be obtained by creating a 'constructed week' (Jones and Carter, 1959), using a procedure established by Stempel (1952), who examined the representativeness of samples of six, twelve, eighteen, twenty-four and forty-eight issues of a newspaper and concluded that 'increasing the sample size beyond twelve does not produce marked differences in the results' (Stempel, 1952: 333). Samples 'were selected by making a random start and taking every Nth issue for the entire year. The size of the gap was chosen so that the day of the week moved up one with each choice' (Stempel, 1952: 333). This principle of randomness was applied in the construction of this corpus, so that for each selected newspaper a different seven-day week was established, with a week consisting of six issues, each for a different day of the week, plus one Sunday for which no issue was included. The same constructed week was applied to each indicator year from 1880–2000, to ensure comparability of data. Because of the changing size of newspapers, the files created by this method were not of similar size (see Table 2). However, as this data was

being used primarily to establish frequency and distribution patterns for words of Maori origin in New Zealand English, the unequal sizes of the files did not undermine the robustness of the findings[3]. The one exception to this method was for the indicator year 1850. This was necessary because:

- newspapers were not published six days a week, thus the key denominator of six could not be applied; and,
- the amount of content that fitted the working definition of New Zealand English was not large; an examination of three months' publication of the *Wellington Independent* for 1850 yielded a useable file of a little over 100,000 words.

Therefore, to ensure a sufficiently large body of data for 1850, the entire year of publication was analysed rather than endeavouring to create a representative sample.

2.3 The need for other data sources

Whilst newspapers form the backbone of this corpus, there was a danger that data obtained from just one source might not prove representative of written New Zealand English. It was necessary, therefore, to expand the findings with data from other sources. This was also necessary because the words of Maori origin in New Zealand English are, by and large, low frequency words. Kennedy and Yamazaki (1999) and Macalister (1999) found that words of Maori origin account for roughly six words per 1,000 in written New Zealand English. Furthermore, of those words, around two-thirds can be expected to be proper nouns of one sort or another. Indeed, as newspaper stories focus on the facts of an event, proper nouns are likely to have a greater presence in the media than in other genres. Thus, while newspapers could be expected to provide reliable information about diachronic variation in frequency and semantic categories, they may not have supplied a sufficiently rich range of vocabulary, nor examples of the vocabulary in use, to investigate satisfactorily other questions. Analysing language use in other genres would also allow triangulation of the findings from newspapers.

A further reason for using additional data sources was to circumvent any possible constraints on the use or treatment of words of Maori origin in newspapers as a result of the dictates of style books. In the event, this appeared to have been, although a valid concern, not an actual threat. Three of the four newspapers in this study responded to a query

---

[3] Similarly, common-sense suggests that the actual word types would have differed to some extent, if a different set of dates had been used for the 'constructed weeks', but the trends would be expected to be similar.

about their approach to the treatment of Maori-origin words, and all three indicated that it was a matter of individual judgment as to which words were translated (on first mention) and which were not.

Parliamentary Debates

The second source selected were parliamentary debates, as these closely replicated the newspaper files in time span and indicator year, thus allowing the opportunity to triangulate the findings for trends in use of words of Maori origin. The first New Zealand Parliament met in Auckland in 1854, following elections the previous year. The *New Zealand Parliamentary Debates* for 1854–1866 were compiled *post facto*, from newspaper reports and other sources (including the speakers' own records)[4]. While they may not, as a result, be entirely accurate, they do represent a valuable, and available, record of language use in New Zealand at that time. As such, they provide a source of comparable data with newspapers for the period under investigation.

A preliminary examination of *Hansard* (as the *NZPD* are usually called) suggested that the use of words of Maori origin was virtually non-existent in debates on what may be termed "general" bills as opposed to "Maori-specific" bills. The data collection, therefore, focussed on debates on "Maori-specific" issues, identified through the indices of the relevant volumes of *Hansard.* As closely as possible, these smaller files repeated the indicator years used for the newspapers. However, as the years 1940 and 1970 produced a very small body of data (see Table 2), the debates of the preceding year were also incorporated. This, in effect, replicated the practice for the third main data source, discussed below.

Finally, it should be noted that although the New Zealand parliament was bi-cameral until the abolition of the (nominated) Legislative Council on 1 January 1951, to ensure comparability these files consisted of information from the (elected) House of Representatives only.

School Journals

The third source of data included in the corpus came from the *School Journals*, an uncontrolled graded reader published for the Ministry of Education in four parts with each part targeting a different age group. Part One, for example, targets seven to eight year olds, while Part Four is written for eleven to thirteen year olds. Each part appears in a number of issues each year, ranging from three to five. The *School Journals* contain

---

[4] As the first New Zealand parliament did not meet until 1854, there was no reduplication of material contained in the newspaper file for 1850. From 1867 parliamentary staff recorded speeches made in the House, and so *Hansard* was not dependent on newspaper reports in any of the indicator years.

informative and imaginative prose, as well as poetry and plays[5]. They have been published since 1907[6] and, as a result, could be used to match the newspaper and *Hansard* files for the twentieth-century indicator years and therefore to triangulate the findings for trends in the use of words of Maori origin. Further, as a genre of writing for young people, they provided a data source that was not available elsewhere (see, for example, Bauer, 1993), and as such had the potential to enrich, and broaden, the range of data gained from the other files. Consideration was, of course, given to possible objections to their use. One such may be that because they represent just one genre, (that is, writing for young people), and even though they are authentic, unsimplified texts, they cannot be taken as representative of New Zealand English. However, earlier research which compared a 1990s *School Journals* corpus with Kennedy and Yamazaki's analysis of the *Wellington Corpus of Written New Zealand English* (Macalister, 1999) established that the *Journals* were a robust representation of New Zealand English of the 1990s, at least in terms of the frequency of use of words of Maori origin, and the patterns of use of those words. The assumption was that the same would hold true for the *Journals* of the 1900s, the 1930s, and the 1960s although there are no studies equivalent to Kennedy and Yamazaki's for those periods against which this assumption can be tested.

A further possible objection relating to the *School Journals* was the fact that the writing is the work of professional, and sometimes prominent, writers. Sinclair (1991: 16–17) has warned against the temptation to over-represent literary language in a corpus because 'well-known writers tend to have unusual ways of writing'. However, presumably because the target audience of the *Journals* is young readers, the writing is not noticeably stylised or innovative.

On balance, the *School Journals* were considered suitable for discerning trends in New Zealand English for a number of reasons. One of their goals is 'to help children understand themselves, their families, and the wider social and natural environment in which they live' (*Guidelines for Contributors*). Thus, the *Journal*, as Beaglehole (1982: 38) points out, 'has always been a potential influence in formulating and reinforcing the self-images and cultural attitudes of those who read it' because of its ubiquitous presence and use in schools. Furthermore, as the social and natural environment has changed, so has the *Journal*. In various ways, therefore, the *Journal* encapsulates the evolution of New Zealand since 1907. Nor is it the subject matter of the *Journal* alone that conveys information about the evolution of New Zealand. As the *Journal* is a

---

[5] Any reader interested in further information can go to:
http://www.learningmedia.co.nz/nz/online/authorsartists/seriesguidelines/schooljournal#abouthteschooljournal
[6] During this period they have been officially called *The School Journal, The New Zealand School Journal, New Zealand School Journal,* and are currently *School Journal*. The *School Journal* is supplied free to all schools in New Zealand.

written medium which has been a potential influence in shaping and reinforcing notions of language use, it is also a resource for measuring trends in New Zealand English. Furthermore, as the *Journal* is published for New Zealand children the assumption is that the language used is familiar to and understood by the readers, that it is, in fact, representative of New Zealand English at the time of publication. From corpus analysis of the *Journal*, therefore, it would be possible to identify changes in the frequency, the patterns of use, and the treatment of Maori words in New Zealand English.

2.4 Data storage and approach to data analysis

Practical considerations such as available time and likely benefit to other researchers meant that the analysis of the newspapers for five of the six indicator years was undertaken using pre-electronic methods, that is, a manual search of the newspapers was undertaken. This entailed:

- identification and categorisation of Maori word tokens;
- counting the numbers of words in numerous ten-line blocks to establish an average number of words per line; and,
- counting the numbers of lines per column

With any search undertaken in such a way, there arise, inevitably, questions of reliability because of the potential for human error. To ascertain the degree of variance likely to occur, three constructed weeks for 1880 were re-visited a year after the initial investigation. As Table 1 shows, there was a high degree of consistency between the results for the two searches.

	Number Maori word tokens		Estimated file size		Maori words per 1,000 tokens	
	1st Search	2nd Search	1st Search	2nd Search	1st Search	2nd Search
NZH	839	840	107,897	108,452	7.70	7.70
NZTimes	224	216	66,924	66,144	3.30	3.26
ODT	402	408	75,501	73,339	5.30	5.56
Overall	1,465	1,464	250,322	247,935	5.85	5.90

Table 1: Reliability of manual search of newspapers

The number of Maori word tokens found was almost the same, with a difference of one. There were larger differences with the number of Maori word tokens for the individual newspapers, but not enough to affect in any serious way the number of words of Maori origin found to be occurring per 1,000 tokens. Also, the estimated total numbers of words in the constructed weeks for the *New Zealand Herald* and the *New Zealand Times* were within a few hundred of each other. The difference between the estimated totals for the *Otago Daily Times* week was larger – over two thousand – which may have been caused by that newspaper's use of a wider range of font sizes, which in turn affected the average numbers of words found per line and lines per page. Even so, the impact on the occurrence of words of Maori origin per 1,000 tokens was not of serious concern.

For the year 2000, the final indicator year, the methodology was altered slightly as it was possible to create a computer-readable file for the four newspapers. This was a labour-intensive process that needed to be approached in three different ways.

1. The *Dominion* sent a "dump" of the entire edition for each day of publication in the constructed week. From this "dump", international stories and any other content that did not fit the parameters established for inclusion were edited out. In addition, as can be seen in Figure 1, every story was prefaced by extraneous information that also needed to be removed. Figure 2 shows the information that remained post-editing, and was, therefore, included in the *Dominion* file for the year 2000.

2. Both the *Otago Daily Times* and the *New Zealand Herald* have web-sites on which substantial portions of the daily edition are posted. Working from the hard copies of editions available in Wellington[7], stories were copied and pasted. Stories that were not available on-line were later photocopied from the paper edition and scanned into the database.

3. The *Wanganui Chronicle* was not available in electronic form. Therefore, each entire edition had to be photocopied, cut up, and scanned[8]. Manual correction of the text was also necessary, as the OCR programme did not identify a wide range of misreadings. The misreadings were actual words (e.g., *tip* for *up*), and thus auto-replacement was not possible.

---

[7] Editions available in Wellington may have been slightly different from the metropolitan editions available in Auckland and Dunedin, just as an early edition may have differed a little from a late edition. However, this seemed to be unavoidable and had to be accepted.
[8] As an indication of the time involved, it took twelve hours, on average, for one edition of the *Wanganui Chronicle* to be converted to an electronically readable form.

Publication: DOM	**Date:** 17 Nov 2000	**Page:** 1

Headline: Victorious IRD battlers tell others to fight on

Author: ALLEY Oskar

Subjects: GOVERNMENT DEPARTMENTS;TAXATION ( )

Section: NEWS **Sub Section:** NATIONAL

Caption(s): JOHN SELKIRK;What a difference a day makes: the Willises after the
compensation news

Edition: 2 **Words:** 555 **Source:** **Copyright:**
Flags:

Figure 1: Sample of information heading a *Dominion* article

Victorious IRD battlers tell others to fight on

What a difference a day makes: the Willises after the compensation news

Figure 2: Sample of information retained post-editing

It was tempting to correct typographical errors, such as *innacurate* (for *inaccurate*) and *fods* (for *foods*), and to impose uniform orthographical practices, particularly as relates to the use of hyphens and apostrophes. At times these were errors, for example *lets* for *let's*, but on others the differences were legitimate variations, for example *selfconscious* or *self-conscious, email* versus *e-mail*. However, the temptation was resisted and the integrity of the texts was retained.

Once the newspapers had been gathered as daily files, each individual story was then assigned to one of seven sub-genres (New Zealand news, politics, business, sport, features and columns, editorial, reviews), so that it is possible to access the 2000 newspaper file in three ways: in its entirety as a representation of written New Zealand English, by regional sub-file, or by specific sub-genre (Macalister, 2001).

As with the newspapers, the analysis for five of the six indicator years of *Hansard* was undertaken using pre-electronic methodology. For the sixth indicator year, the year 2000, an electronic "dump" of each sitting day's proceedings was received, totalling over five-million word tokens prior to editing.

For the *School Journal* files, however, all were stored electronically. For the earliest indicator years this entailed scanning, as with the *Wanganui Chronicle*, but for the year 2000 the texts were received electronically from the publisher. In the end, four chronologically distinct files for the *School Journals* were created, with two years of publication being included in each file to ensure that they were large enough for their intended purpose. Even so, the *School Journals* files are relatively small (see Table 2) with the combined total being about one million words. Not all content was included, however. The principal exclusions were covers, contents page, acknowledgements inside back cover, any item written wholly in Maori rather than English, and extended stretches of Maori, such as waiata (or songs), where Maori words were clearly not being used as part of New Zealand English[9]. The files were categorised by level, the *Journal* being published in parts, each with a different target age group, and by genre, with a particular distinction being made between imaginative and informative prose.

3. Conclusion

This corpus consists of running text from three sources, which produced sixteen files and a corpus of a little under five-and-a-half million words, as shown in Table 2.

	1850	1880	1910	1940	1970	2000	Total
News-papers	355,074	596,891	661,891	671,043	668,755	819,495	3,773,149
Hansard	9,188	260,346	101,618	55,386	90,627	80,212	597,377
School Journal	0	0	299,460	287,127	327,734	137,699	1,052,020
Total	364,262	857,237	1,062,969	1,013,556	1,087,116	1,037,406	**5,422,546**

Table 2: Description of the corpus

At almost five-and-a-half million words, this was the largest yet assembled for the study of New Zealand English. It was designed to be representative of standard New Zealand English in general use and constructed in such a way as to allow meaningful comparison of language in use at different

---

[9] The rule of thumb here and elsewhere was that, to be included, Maori language lexical units needed to be potentially understandable to non-Maori-language speakers. In essence, therefore, the formal distinction between code-switching (inter-sentential language change) and code-mixing (intra-sentential language change) was maintained, with the former being ruled out and the latter generally accepted. Note, however, that the switching/mixing distinction is not universally accepted. Myers-Scotton (1989: 334), for instance, labels the distinction 'poorly motivated'.

times over a 150-year period. Analysis of the corpus provided empirical support for earlier claims about the presence of Maori words in New Zealand English (Macalister, 2006). The electronically-readable corpus files have been made available to other researchers, and the expectation is that this will provide material for fresh insights into aspects of New Zealand English in the future. In addition, this is envisaged as an on-going project. The intention is to make increasing amounts of the corpus electronically-readable, once funding is available, and to add a seventh indicator year to the corpus in 2010.

References

Bauer, L. 1993. Manual of Information to Accompany the Wellington Corpus of New Zealand English. Wellington: Department of Linguistics, Victoria University of Wellington.

Beaglehole, D. 1982. 'The Maori in the *School Journal*: 1907–1981', Education 31 (1), pp. 38-41.

Bell, A. 1991. The Language of News Media. Oxford: Blackwell.

Deverson, T. 1991. 'New Zealand English lexis: the Maori dimension', English Today 26, pp. 18-25.

Gordon, E., Campbell, L., Hay, J., Maclagan, M. and P. Trudgill. 2004. New Zealand English: Its Origins and Evolution. Cambridge: Cambridge University Press.

Grey, A. 1994. Aotearoa and New Zealand: A Historical Geography. Christchurch: Canterbury University Press.

Holmes, J. 1995. 'The Wellington Corpus of Spoken New Zealand English: A Progress Report', New Zealand English Newsletter 9, pp. 5-8.

Holmes, J., Vine, B. and G. Johnson. 1998. Guide to the Wellington Corpus of Spoken New Zealand English. Wellington: Victoria University of Wellington.

Jones, R.L. and R.E. Carter, Jr. 1959. 'Some Procedures for Estimating 'News Hole' in Content Analysis', Public Opinion Quarterly 23 (3), pp. 399-403.

Kennedy, G. 2001. 'Lexical borrowing from Maori in New Zealand English' in B. Moore (ed.) Who's Centric Now? The Present State of Post-Colonial Englishes, pp. 59–81. Melbourne: Oxford University Press.

Kennedy, G. and S. Yamazaki. 1999. 'The Influence of Maori on the New Zealand English Lexicon' in J.M. Kirk (ed.) Corpora Galore: Analyses and Techniques in Describing English, pp. 33-44. Amsterdam: Rodopi.

Macalister, J. 1999. 'Trends in New Zealand English: Some Observations on the Presence of Maori Words in the Lexicon', New Zealand English Journal 13, pp. 38-49.

Macalister, J. 2001. 'Introducing a New Zealand Newspaper Corpus', New Zealand English Journal 15, pp. 35-41.

Macalister, J. 2006. 'The Maori presence in the New Zealand English lexicon, 1850-2000: Evidence form a corpus-based study', English World-Wide 27 (1), pp. 1-24.

McBurney, S. 1887. 'Colonial Pronunciation', *The Press*, 5 October 1887.

Morris, E. 1898. Austral English. London: Macmillan

Myers-Scotton, C. 1989. 'Codeswitching with English: types of switching, types of communities', World Englishes 8 (3), pp. 333-46.

Scholefield, G.H. 1958. Newspapers in New Zealand. Wellington: A.H. and A.W. Reed.

Sinclair, J. 1991. Corpus Concordance Collocation. Oxford: Oxford University Press.

Stempel, G.H. III. 1952. 'Sample size for classifying subject matter in dailies', Journalism Quarterly 29 (3), pp. 333–34.