
A New Academic Word List

AVERIL COXHEAD

*Victoria University of Wellington
Wellington, New Zealand*

This article describes the development and evaluation of a new academic word list (Coxhead, 1998), which was compiled from a corpus of 3.5 million running words of written academic text by examining the range and frequency of words outside the first 2,000 most frequently occurring words of English, as described by West (1953). The AWL contains 570 word families that account for approximately 10.0% of the total words (tokens) in academic texts but only 1.4% of the total words in a fiction collection of the same size. This difference in coverage provides evidence that the list contains predominantly academic words. By highlighting the words that university students meet in a wide range of academic texts, the AWL shows learners with academic goals which words are most worth studying. The list also provides a useful basis for further research into the nature of academic vocabulary.

One of the most challenging aspects of vocabulary learning and teaching in English for academic purposes (EAP) programmes is making principled decisions about which words are worth focusing on during valuable class and independent study time. Academic vocabulary causes a great deal of difficulty for learners (Cohen, Glasman, Rosenbaum-Cohen, Ferrara, & Fine, 1988) because students are generally not as familiar with it as they are with technical vocabulary in their own fields and because academic lexical items occur with lower frequency than general-service vocabulary items do (Worthington & Nation, 1996; Xue & Nation, 1984).

The General Service List (GSL) (West, 1953), developed from a corpus of 5 million words with the needs of ESL/EFL learners in mind, contains the most widely useful 2,000 word families in English. West used a variety of criteria to select these words, including frequency, ease of learning, coverage of useful concepts, and stylistic level (pp. ix-x). The GSL has been criticised for its size (Engels, 1968), age (Richards, 1974), and need for revision (Hwang, 1989). Despite these criticisms, the GSL covers up to 90% of fiction texts (Hirsh, 1993), up to 75% of nonfiction texts (Hwang, 1989), and up to 76% of the Academic Corpus (Coxhead,

1998), the corpus of written academic English compiled for this study. There has been no comparable replacement for the GSL up to now.

Academic words (e.g., *substitute, underlie, establish, inherent*) are not highly salient in academic texts, as they are supportive of but not central to the topics of the texts in which they occur. A variety of word lists have been compiled either by hand or by computer to identify the most useful words in an academic vocabulary. Champion and Elley (1971) and Praninskas (1972) based their lists on corpora and identified words that occurred across a range of texts whereas Lynn (1973) and Ghadessy (1979) compiled word lists by tracking student annotations above words in textbooks. All four studies were developed without the help of computers. Xue and Nation (1984) created the University Word List (UWL) by editing and combining the four lists mentioned above. The UWL has been widely used by learners, teachers, course designers, and researchers. However, as an amalgam of the four different studies, it lacked consistent selection principles and had many of the weaknesses of the prior work. The corpora on which the studies were based were small and did not contain a wide and balanced range of topics.

An academic word list should play a crucial role in setting vocabulary goals for language courses, guiding learners in their independent study, and informing course and material designers in selecting texts and developing learning activities. However, given the problems with currently available academic vocabulary lists, there is a need for a new academic word list based on data gathered from a large, well-designed corpus of academic English. The ideal word list would be divided into smaller, frequency-based sublists to aid in the sequencing of teaching and in materials development. A word list based on the occurrence of word families in a corpus of texts representing a variety of academic registers can provide information about how words are actually used (Biber, Conrad, & Reppen, 1994).

The research reported in this article drew upon principles from corpus linguistics (Biber, Conrad, & Reppen, 1998; Kennedy, 1998) to develop and evaluate a new academic word list. After discussing issues that arise in the creation of a word list through a corpus-based study, I describe the methods used in compiling the Academic Corpus and in developing the AWL. The next section examines the coverage of the AWL relative to the complete Academic Corpus and to its four discipline-specific subcorpora. To evaluate the AWL, I discuss its coverage of (a) the Academic Corpus along with the GSL (West, 1953), (b) a second collection of academic texts, and (c) a collection of fiction texts, and compare it with the UWL (Xue & Nation, 1984). In concluding, I discuss the list's implications for teaching and for materials and course design, and I outline future research needs.

THE DEVELOPMENT OF ACADEMIC CORPORA AND WORD LISTS

Teachers and materials developers who work with vocabulary lists often assume that frequently occurring words and those which occur in many different kinds of texts may be more useful for language learners to study than infrequently occurring words and those whose occurrences are largely restricted to a particular text or type of text (Nation, in press; West, 1953). Given the assumption that frequency and coverage are important criteria for selecting vocabulary, a corpus, or collection of texts, is a valuable source of empirical information that can be used to examine the language in depth (Biber, Conrad, & Reppen, 1994). However, exactly how a corpus should be developed is not clear cut. Issues that arise include the representativeness of the texts of interest to the researcher (Biber, 1993), the organization of the corpus, its size (Biber, 1993; Sinclair, 1991), and the criteria used for word selection.

Representation

Research in corpus linguistics (Biber, 1989) has shown that the linguistic features of texts differ across registers. Perhaps the most notable of these features is vocabulary. To describe the vocabulary of a particular register, such as academic texts, the corpus must therefore contain texts that are representative of the varieties of texts they are intended to reflect (Atkins, Clear, & Ostler, 1992; Biber, 1993; Sinclair, 1991). Sinclair (1991) warns that a corpus should contain texts whose sizes and shapes accurately reflect the texts they represent. If long texts are included in a corpus, "peculiarities of an individual style or topic occasionally show through" (p. 19), particularly through the vocabulary. Making use of a variety of short texts allows more variation in vocabulary (Sutarsyah, Nation, & Kennedy, 1994). Inclusion of texts written by a variety of writers helps neutralise bias that may result from the idiosyncratic style of one writer (Atkins et al., 1992; Sinclair, 1991) and increases the number of lexical items in the corpus (Sutarsyah et al., 1994).

Scholars who have compiled corpora have attempted to include a variety of academic texts. Campion and Elley's (1971) corpus consisted of 23 textbooks, 19 lectures published in journals, and a selection of university examination papers. Praninskas (1972) used a corpus of 10 first-year, university-level arts and sciences textbooks that were required reading at the American University of Beirut. Lynn (1973) and Ghadessy (1979) both focussed on textbooks used in their universities. Lynn's corpus included 52 textbooks and 4 classroom handouts from 50

students of accounting, business administration, and economics from which 10,000 annotations were collected by hand. The resulting list contained 197 word families arranged from those occurring the most frequently (39 times) to those occurring the least frequently. Words occurring fewer than 10 times were omitted from the list (p. 26). Ghadessy compiled a corpus of 20 textbooks from three disciplines (chemistry, biology, and physics). Words that students had glossed were recorded by hand, and the final list of 795 items was then arranged in alphabetical order (p. 27). Relative to this prior work, the corpus compiled for the present study considerably expands the representation of academic writing in part by including a variety of academic sources besides textbooks.

Organization

A register such as academic texts encompasses a variety of subregisters. An academic word list should contain an even-handed selection of words that appear across the various subject areas covered by the texts contained within the corpus. Organizing the corpus into coherent sections of equal size allows the researcher to measure the range of occurrence of the academic vocabulary across the different disciplines and subject areas of the corpus. Champion and Elley (1971) created a corpus with 19 academic subject areas, selecting words occurring outside of the first 5,000 words of Thorndike and Lorge's (1944) list and excluding words encountered in only one discipline (p. 7). The corpus for the present study involved 28 subject areas organised into 7 general areas within each of four disciplines: arts, commerce, law, and science.

Size

A corpus designed for the study of academic vocabulary should be large enough to ensure a reasonable number of occurrences of academic words. According to Sinclair (1991), a corpus should include millions of running words (tokens) to ensure that a very large sample of language is available (p. 18).¹ The exact amount of language required, of course, depends on the purpose and use of the research; however, in general more language means that more information can be gathered about lexical items and more words in context can be examined in depth.

¹ The term *running words* (or *tokens*) refers to the total number of word forms in a text, whereas the term *individual words* (*types*) refers to each different word in a text, irrespective of how many times it occurs.

In the past, researchers attempted to work with academic corpora by hand, which limited the numbers of words they could analyze. Campion and Elley (1971), in their corpus of 301,800 running words, analysed 234,000 words in textbooks, 57,000 words from articles in journals, and 10,800 words in a number of examination papers (p. 4). Praninskas's (1972) corpus consisted of approximately 272,000 running words (p. 8), Lynn (1973) examined 52 books and 4 classroom handouts (p. 26), and Ghadessy (1979) compiled a corpus of 478,700 running words. Praninskas (1972) included a criterion of range in her list and selected words that were outside the GSL (West, 1953).

In the current study, the original target was to gather 4.0 million words; however, time pressures and lack of available texts limited the corpus to approximately 3.5 million running words. The decision about size was based on an arbitrary criterion relating to the number of occurrences necessary to qualify a word for inclusion in the word list: If the corpus contained at least 100 occurrences of a word family, allowing on average at least 25 occurrences in each of the four sections of the corpus, the word was included. Study of data from the Brown Corpus (Francis & Kucera, 1982) indicated that a corpus of around 3.5 million words would be needed to identify 100 occurrences of a word family.

Word Selection

An important issue in the development of word lists is the criteria for word selection, as different criteria can lead to different results. Researchers have used two methods of selection for academic word lists. As mentioned, Lynn (1973) and Ghadessy (1979) selected words that learners had annotated regularly in their textbooks, believing that the annotation signalled difficulty in learning or understanding those words during reading. Campion and Elley (1971) selected words based on their occurrence in 3 or more of 19 subject areas and then applied criteria, including the degree of familiarity to native speakers. However, the number of running words in the complete corpus was too small for many words to meet the initial criterion. Praninskas (1972) also included a criterion of range in her list; however, the range of subject areas and number of running words was also small, resulting in a small list without much variety in the words.

Another issue that arises in developing word lists is defining what to count as a word. The problem is that lexical items that may be morphologically distinct from one another are, in fact, strongly enough related that they should be considered to represent a single lexical item. To address this issue, word lists for learners of English generally group words into families (West, 1953; Xue & Nation, 1984). This solution is

supported by evidence suggesting that word families are an important unit in the mental lexicon (Nagy, Anderson, Schommer, Scott, & Stallman, 1989, p. 262). Comprehending regularly inflected or derived members of a family does not require much more effort by learners if they know the base word and if they have control of basic word-building processes (Bauer & Nation, 1993, p. 253). In the present study, therefore, words were defined through the unit of the word family, as illustrated in Table 1.

For the creation of the AWL, a word family was defined as a stem plus all closely related affixed forms, as defined by Level 6 of Bauer and Nation's (1993) scale. The Level 6 definition of affix includes all inflections and the most frequent, productive, and regular prefixes and suffixes (p. 255). It includes only affixes that can be added to stems that can stand as free forms (e.g., *specify* and *special* are not in the same word family because *spec* is not a free form).

Research Questions

The purpose of the research described here was to develop and evaluate a new academic word list on the basis of a larger, more principled corpus than had been used in previous research. Two questions framed the description of the AWL:

1. Which lexical items occur frequently and uniformly across a wide range of academic material but are not among the first 2,000 words of English as given in the GSL (West, 1953)?
2. Do the lexical items occur with different frequencies in arts, commerce, law, and science texts?

TABLE 1
Sample Word Families From the Academic Word List

| <i>concept</i> | legislate | <i>indicate</i> |
|-------------------|--------------------|-----------------|
| conception | legislated | indicated |
| concepts | legislates | indicates |
| conceptual | legislating | indicating |
| conceptualisation | <i>legislation</i> | indication |
| conceptualise | legislative | indications |
| conceptualised | legislator | indicative |
| conceptualises | legislators | indicator |
| conceptualising | legislature | indicators |
| conceptually | | |

Note. Words in italics are the most frequent form in that family occurring in the Academic Corpus.

The evaluation of the AWL considered the following questions:

3. What percentage of the words in the Academic Corpus does the AWL cover?
4. Do the lexical items identified occur frequently in an independent collection of academic texts?
5. How frequently do the words in the AWL occur in nonacademic texts?
6. How does the AWL compare with the UWL (Xue & Nation, 1984)?

METHODOLOGY

The development phase of the project identified words that met the criteria for inclusion in the AWL (Research Questions 1 and 2). In the evaluation phase, I calculated the AWL's coverage of the original corpus and compared the AWL with words found in another academic corpus, with those in a nonacademic corpus, and with another academic word list (Questions 3–6).

Developing the Academic Corpus

Developing the corpus involved collecting each text in electronic form, removing its bibliography, and counting its words. After balancing the number of short, medium-length, and long texts (see below for a discussion on the length of texts), each text was inserted into its subject-area computer file in alphabetical order according to the author's name. Each subject-area file was then inserted into a discipline master file, in alphabetical order according to the subject. Any text that met the selection criteria but was not included in the Academic Corpus because its corresponding subject area was complete was kept aside for use in a second corpus used to test the AWL's coverage at a later stage. The resulting corpus contained 414 academic texts by more than 400 authors, containing 3,513,330 tokens (running words) and 70,377 types (individual words) in approximately 11,666 pages of text. The corpus was divided into four subcorpora: arts, commerce, law, and science, each containing approximately 875,000 running words and each subdivided into seven subject areas (see Table 2).

The corpus includes the following representative texts from the academic domain: 158 articles from academic journals, 51 edited academic journal articles from the World Wide Web, 43 complete university textbooks or course books, 42 texts from the Learned and Scientific section of the Wellington Corpus of Written English (Bauer, 1993), 41

TABLE 2
Composition of the Academic Corpus

| | Discipline | | | | |
|---------------|-------------|---------------|---------------------|------------------|---------|
| | Arts | Commerce | Law | Science | Total |
| Running words | 883,214 | 879,547 | 874,723 | 875,846 | 351,333 |
| Texts | 122 | 107 | 72 | 113 | 414 |
| Subject areas | Education | Accounting | Constitutional | Biology | |
| | History | Economics | Criminal | Chemistry | |
| | Linguistics | Finance | Family and | Computer science | |
| | Philosophy | Industrial | medicolegal | Geography | |
| | Politics | relations | International | Geology | |
| | Psychology | Management | Pure commercial | Mathematics | |
| | Sociology | Marketing | Quasi-commercial | Physics | |
| | | Public policy | Rights and remedies | | |

texts from the Learned and Scientific section of the Brown Corpus (Francis & Kucera, 1982), 33 chapters from university textbooks, 31 texts from the Learned and Scientific section of the Lancaster-Oslo/Bergen (LOB) Corpus (Johansson, 1978), 13 books from the Academic Texts section of the MicroConcord academic corpus (Murison-Bowie, 1993), and 2 university psychology laboratory manuals.

The majority of the texts were written for an international audience. Sixty-four percent were sourced in New Zealand, 20% in Britain, 13% in the United States, 2% in Canada, and 1% in Australia. It is difficult to say exactly what influence the origin of the texts would have on the corpus, for even though a text was published in one country, at least some of the authors may well have come from another.

The Academic Corpus was organized to allow the range of occurrence of particular words to be examined. Psychology and sociology texts were placed in the arts section on the basis of Biber's (1989) finding that texts from the social sciences (psychology and sociology) shared syntactic characteristics with texts from the arts (p. 28). Lexical items may well pattern similarly. Placing the social science subject areas in the science section of the Academic Corpus might have introduced a bias: The psychology and sociology texts might have added lexical items that do not occur in any great number in any other subject in the science section. The presence of these items, in turn, would have suggested that science and arts texts share more academic vocabulary items than is generally true.

With the exception of the small number of texts from the Brown (Francis & Kucera, 1982), LOB (Johansson, 1978), and Wellington

(Bauer, 1993) corpora, the texts in the Academic Corpus were complete. The fact that frequency of occurrence of words was only one of the criteria for selecting texts minimized any possible bias from word repetition within longer texts. To maintain a balance of long and short texts, the four main sections (and, within each section, the seven subject areas) each contained approximately equal numbers of short texts (2,000–5,000 running words), medium texts (5,000–10,000 running words), and long texts (more than 10,000 running words). The breakdown of texts in the four main sections was as follows: arts—18 long, 35 medium; commerce—18 long, 37 medium; law—23 long, 22 medium; and science—19 long, 37 medium.

Developing the Academic Word List

The corpus analysis programme Range (Heatley & Nation, 1996) was used to count and sort the words in the Academic Corpus. This programme counts the frequency of words in up to 32 files at a time and records the number of files in which each word occurs (range) and the frequency of occurrence of the words in total and in each file.

Words were selected for the AWL based on three criteria:

1. *Specialised occurrence*: The word families included had to be outside the first 2,000 most frequently occurring words of English, as represented by West's (1953) GSL.
2. *Range*: A member of a word family had to occur at least 10 times in each of the four main sections of the corpus and in 15 or more of the 28 subject areas.
3. *Frequency*: Members of a word family had to occur at least 100 times in the Academic Corpus.

Frequency was considered secondary to range because a word count based mainly on frequency would have been biased by longer texts and topic-related words. For example, the *Collins COBUILD Dictionary* (1995) highlights *Yemeni* and *Lithuanian* as high-frequency words, probably because the corpus on which the dictionary is based contains a large number of newspapers from the early 1990s.

The conservative threshold of a frequency of 100 was applied strictly for multiple-member word families but not so stringently for word families with only one member, as single-member families operate at a disadvantage in gaining a high frequency of occurrence. In the Academic Corpus, the word family with only one member that occurs the least frequently is *forthcoming* (80 occurrences).

RESULTS

Description

Occurrence of Academic Words

The first research question asked which lexical items beyond the first 2,000 in West's (1953) GSL occur frequently across a range of academic texts. In the Academic Corpus, 570 word families met the criteria for inclusion in the AWL (see Appendix A). Some of the most frequent word families in the AWL are *analyse*, *concept*, *data*, and *research*. Some of the least frequent are *convince*, *notwithstanding*, *ongoing*, *persist*, and *whereby*.

Differences in Occurrence of Words Across Disciplines

The second question was whether the lexical items selected for the AWL occur with different frequencies in arts, commerce, law, and science texts. The list appears to be slightly advantageous for commerce students, as it covers 12.0% of the commerce subcorpus. The coverage of arts and of law is very similar (9.3% and 9.4%, respectively), and the coverage of science is the lowest among the four disciplines (9.1%). The 3.0% difference between the coverage of the commerce subcorpus and the coverage of the other three subcorpora may result from the presence of key lexical items such as *economic*, *export*, *finance*, and *income*, which occur with very high frequency in commerce texts. (See Appendix B for excerpts from texts in each section of the Academic Corpus.)

The words in the AWL occur in a wide range of the subject areas in the Academic Corpus. Of the 570 word families in the list, 172 occur in all 28 subject areas, and 263 (172 + 91) occur in 27 or more subject areas (see Table 3). In total, 67% of the word families in the AWL occur in 25 or more of the 28 subject areas, and 94% occur in 20 or more.

Evaluation

Coverage of the Academic Corpus Beyond the GSL

The AWL accounts for 10.0% of the tokens in the Academic Corpus. This coverage is more than twice that of the third 1,000 most frequent words, according to Francis and Kucera's (1982) count, which cover 4.3% of the Brown Corpus. Taken together, the first 2,000 words in West's (1953) GSL and the word families in the AWL account for approximately 86% of the Academic Corpus (see Table 4). Note that the AWL's coverage of the Academic Corpus is double that of the second

TABLE 3

Subject-Area Coverage of Word Families in the Academic Word List

| No. of word families | Subject areas in which they occurred | No. of word families | Subject areas in which they occurred |
|----------------------|--------------------------------------|----------------------|--------------------------------------|
| 172 | 28 | 20 | 21 |
| 91 | 27 | 15 | 20 |
| 58 | 26 | 9 | 19 |
| 62 | 25 | 9 | 18 |
| 43 | 24 | 5 | 17 |
| 43 | 23 | 5 | 16 |
| 33 | 22 | 4 | 15 |

Note. Total subject areas = 28; total word families = 570.

1,000 words of the GSL. The AWL and the GSL combined have a total of 2,550 word families, and all but 12 of those in the GSL occur in the Academic Corpus.

The AWL, the first 1,000 words of the GSL (West, 1953), and the second 1,000 words of the GSL cover the arts, commerce, and law subcorpora similarly but in very different patterns (see Table 5). The first 1,000 words of the GSL account for fewer of the word families in the commerce subcorpus than in the arts and law subcorpora, but this lower coverage of commerce is balanced by the AWL's higher coverage of this discipline. On the other hand, the AWL's coverage of the arts and law subcorpora is lower than its coverage of the commerce subcorpus, but the GSL's coverage of arts and law is slightly higher than its coverage of commerce. The AWL's coverage of the science subcorpus is 9.1%, which indicates that the list is also extremely useful for science students. The GSL, in contrast, is not quite as useful for science students as it is for arts, commerce, and law students.

TABLE 4

Coverage of the Academic Corpus by the Academic Word List and the General Service List (West, 1953)

| Word list | Coverage of Academic Corpus (%) | No. of word families | |
|----------------------|---------------------------------|----------------------|--------------------|
| | | Total | In Academic Corpus |
| Academic Word List | 10.0 | 570 | 570 |
| General Service List | | | |
| First 1,000 words | 71.4 | 1,001 | 1,000 |
| Second 1,000 words | 4.7 | 979 | 968 |
| Total | 86.1 | 2,550 | 2,538 |

TABLE 5
Coverage of the Four Subcorpora of the Academic Corpus
by the General Service List (West, 1953) and the Academic Word List (%)

| Subcorpus | Academic Word List | General Service List | | Total |
|-----------|--------------------|----------------------|--------------------|-------|
| | | First 1,000 words | Second 1,000 words | |
| Arts | 9.3 | 73.0 | 4.4 | 86.7 |
| Commerce | 12.0 | 71.6 | 5.2 | 88.8 |
| Law | 9.4 | 75.0 | 4.1 | 88.5 |
| Science | 9.1 | 65.7 | 5.0 | 79.8 |

Coverage of Another Academic Corpus

A frequency-based word list that is derived from a particular corpus should be expected to cover that corpus well. The real test is how the list covers a different collection of similar texts. To establish whether the AWL maintains high coverage over academic texts other than those in the Academic Corpus, I compiled a second corpus of academic texts in English, using the same criteria and sources to select texts and dividing them into the same four disciplines. This corpus comprised approximately 678,000 tokens (82,000 in arts, 53,000 in commerce, 143,000 in law, and 400,000 in science) representing 32,539 types of lexical items. This second corpus was made up of texts that had met the criteria for inclusion in the Academic Corpus but were not included either because they were collected too late or because the subject area they belonged to was already complete.

The AWL's coverage of the second corpus is 8.5% (see Table 6), and all 570 word families in the AWL occur in the second corpus. The GSL's coverage of the second corpus (66.2%) is consistent with its coverage of the science section of the Academic Corpus (65.7%). The overall lower coverage of the second corpus by both the AWL and the GSL (79.1%) seems to be partly the result of the large proportion of science texts it contains.

Coverage of Nonacademic Texts

To establish that the AWL is truly an academic word list rather than a general-service word list, I developed a collection of 3,763,733 running words of fiction texts. The collection consisted of 50 texts from Project Gutenberg's (<http://www.gutenberg.net>) collection of texts that were written more than 50 years ago and are thus in the public domain. The

TABLE 6

Coverage of the Academic Corpus and the Second Corpus of Academic Texts by the Academic Word List and the General Service List (West, 1953) (%)

| Word list | Coverage of Academic Corpus | Coverage of second corpus |
|----------------------|-----------------------------|---------------------------|
| Academic Word List | 10.0 | 8.5 |
| General Service List | | |
| First 1,000 words | 71.4 | 66.2 |
| Second 1,000 words | 4.7 | 4.4 |
| Total | 86.1 | 79.1 |

fact that the Academic Corpus contained many more texts (414) is not important, because the central purpose of compiling the fiction collection was to find out whether the AWL families occurred frequently in fiction texts. Neither the number of texts, nor their length, nor the range of lexical items occurring across the texts was crucial to the comparison.

The AWL accounts for approximately 1.4% of the tokens in the fiction collection, much lower than the AWL's 10% coverage of the Academic Corpus. The markedly different coverage suggests that the majority of word families in the AWL are associated particularly with academic writing (see Table 7). The age of the fiction texts may be another reason that the word families in the AWL occur infrequently in the fiction collection, and for words such as *infrastructure*, this is probably true. However, an examination of the AWL words revealed few of this type.

Of the AWL families, 410 (380 + 30) are clearly academic; that is, they occur with much higher frequency in academic than in fiction texts. An additional 86 occur with more than twice the frequency in academic as

TABLE 7

Occurrence of the AWL Word Families in the Academic Corpus and the Fiction Collection

| Frequency of occurrence | No. of AWL word families |
|---|--------------------------|
| Not in fiction collection | 30 |
| In Academic Corpus | |
| Four or more times as frequently as in fiction collection | 380 |
| Three times as frequently as in fiction collection | 34 |
| Twice as frequently as in fiction collection | 52 |
| Less than twice as frequently as in fiction collection | 52 |
| Less frequently than in fiction collection | 22 |
| Total | 570 |

they do in fiction texts. The remaining 74 (those occurring less than twice as frequently in academic texts as they do in fiction texts and those occurring less frequently in academic than in fiction texts) are candidates for inclusion in a new general-service word list, depending on the selection criteria for that list.

Comparison With the University Word List

The UWL (Xue & Nation, 1984), created through the amalgamation of four existing word lists, contains 836 word families consisting of 3,707 types and covers 8.5% of the Learned and Scientific sections of the LOB corpus of written British English (Johansson, 1978) and the parallel Wellington corpus of written English (Bauer, 1993). It covers 9.8% of the Academic Corpus, slightly less than the 10.0% coverage of the corpus by the AWL. Therefore, the AWL, though smaller, gives a better return on learning, as students would need to learn only 570 word families instead of 836 for the same coverage of academic texts.

The overlap between the AWL and the UWL is 51%, with 435 word families occurring in both. This leaves 401 word families occurring only in the UWL and 135 word families occurring only in the AWL. The explanation for the large number of word families occurring in the UWL but not in the AWL lies in the criteria for including word families in the AWL: Members of a word family had to occur at least 100 times in the Academic Corpus. Approximately 150 of the word families that are only in the UWL occurred in the Academic Corpus less than 50 times, or only once in more than 174 pages of 400 words, and therefore would not have been included in the AWL. Other words in the UWL did not meet the range criterion for the AWL.

The UWL contains more than 133 word families that do not occur in all four sections of the Academic Corpus (Table 8). Thus students could learn these words but might rarely or never encounter them in academic texts. Although the UWL contains useful words for students to learn, as shown by the 9.8% coverage of the Academic Corpus, the AWL is smaller, has a higher coverage of academic texts, and covers a far wider range of subject areas.

CONCLUSION

The Academic Word List includes 570 word families that constitute a specialised vocabulary with good coverage of academic texts, regardless of the subject area. It accounts for 10% of the total tokens in the Academic Corpus, and more than 94% of the words in the list occur in 20 or more of the 28 subject areas of the Academic Corpus. These

TABLE 8
Characteristics of the Academic Word List and the
University Word List (Xue & Nation, 1984)

| Characteristic | Academic Word List | University Word List |
|--------------------------------------|--------------------|----------------------|
| Word families (total) | 570 | 836 |
| Types | 3,110 | 3,707 |
| Coverage of Academic Corpus (%) | 10.0 | 9.8 |
| Inclusion of word families | | |
| In four sections of Academic Corpus | 570 | 703 |
| In three sections of Academic Corpus | 0 | 84 |
| In two sections of Academic Corpus | 0 | 39 |
| In one section of Academic Corpus | 0 | 6 |
| Not found in Academic Corpus | 0 | 1 |
| In General Service List (West, 1953) | 0 | 3 |

findings are useful in teaching English and point to directions for future research.

Implications for Teaching

The AWL is the result of a corpus-based study. Such studies create lists, concordances, or data concerning the clustering of linguistic items in coherent, purposeful texts. The use of this research method, however, does not imply that language teaching and learning should rely on decontextualised methods. Instead, the AWL might be used to set vocabulary goals for EAP courses, construct relevant teaching materials, and help students focus on useful vocabulary items.

The AWL will be most valuable in setting goals for EAP courses. This study has identified vocabulary to include in teaching and learning materials, but there remains a need to design tests to diagnose whether learners know this vocabulary and whether attempts to teach and learn it have been successful. Such tests exist for the UWL (Nation, 1983); similar tests based on the AWL are under development.

The UWL and one of its predecessors, the *American University Word List* (Praninskas, 1972), served as the basis for course books specifically designed to teach academic vocabulary (Farid, 1985; Valcourt & Wells, 1999; Yorkey, 1981). It is hoped that authors will undertake to write similar books based on the AWL. In addition, a useful direction for materials development would be the design of texts that provide optimal conditions for meeting and learning academic vocabulary. This initiative might involve adapting academic texts so that the density of unknown

words, particularly academic vocabulary and low-frequency words, is not too high and the opportunities for repeated exposure to the academic vocabulary are optimised.

The direct learning and direct teaching of the words in the AWL also have value. Courses that involve direct attention to language features have been found to result in better learning than courses that rely solely on incidental learning (Ellis, 1990; Long, 1988). Using subdivisions of the AWL, teachers and students can set short-term vocabulary learning goals of reasonable size during courses of study. The AWL can be divided into 10 rank-ordered sublists according to decreasing word family frequency (Table 9). With the exception of Sublist 10, each sublist contains 60 items. The words in the first 3 sublists occur with comparatively high frequency (on average, in at least every 12.0 pages of text). On average, each word in Sublist 1 occurs once in 4.3 pages of academic text, assuming that each page is 400 words long. These 60 words account for more than one third of the total coverage of the list, and the next most frequently occurring 60 words (Sublist 2) provide just half the coverage of the first 60 words. Even though Sublists 5–10 add little to the overall coverage of the AWL, they are worth including, as these less frequent items occur in a wide range of texts and are unlikely to be acquired incidentally through reading.

Direct teaching through vocabulary exercises, teacher explanation, and awareness raising, and deliberate learning using word cards need to be balanced with opportunities to meet the vocabulary in message-focused reading and listening and to use the vocabulary in speaking and writing. For direct study of the vocabulary, teachers and learners can work from the list itself. More than 82% of the words in the AWL are of

TABLE 9
Sublists of the Academic Word List

| Sublist | Items | Coverage of Academic Corpus (%) | Cumulative coverage (%) | Pages per repetition in Academic Corpus |
|---------|-------|---------------------------------|-------------------------|---|
| 1 | 60 | 3.6 | 3.6 | 4.3 |
| 2 | 60 | 1.8 | 5.4 | 8.4 |
| 3 | 60 | 1.2 | 6.6 | 12.3 |
| 4 | 60 | 0.9 | 7.5 | 15.9 |
| 5 | 60 | 0.8 | 8.3 | 19.4 |
| 6 | 60 | 0.6 | 8.9 | 24.0 |
| 7 | 60 | 0.5 | 9.4 | 30.8 |
| 8 | 60 | 0.3 | 9.7 | 49.4 |
| 9 | 60 | 0.2 | 9.9 | 67.3 |
| 10 | 30 | 0.1 | 10.0 | 82.5 |

Greek or Latin origin, indicating that the study of prefixes, suffixes, and stems may be one way to study this vocabulary.

By focusing on this academic vocabulary in both message-focussed and language-focussed ways, learners gain the opportunity to make this important vocabulary a part of their working knowledge of the language and thus help make their academic study more manageable.

Future Research

The results of this research show that the development and use of large corpora hold promise for obtaining information about vocabulary frequency in registers of interest for language teaching. Future research might fruitfully build on these findings in four ways.

1. Compare the findings obtained from the Academic Corpus with those from larger corpora, such as those used for dictionary making. In this study, the collection of texts used for comparison with the Academic Corpus was smaller rather than larger than the original Academic Corpus, and its lack of balance in the number of running words per discipline made a full comparison impossible. In addition, the law subcorpus contained only half the number of short texts (27) as did the other three subcorpora (arts, 60; commerce, 52; science, 57), which may have resulted in less variety in the vocabulary of the law subcorpus. Approximately 6% (or 228,000 running words) of the Academic Corpus consisted of 114 incomplete texts of 2,000 running words that came from the Brown (Francis & Kucera, 1982), LOB (Johansson, 1978), and Wellington (Bauer, 1993) corpora. Whereas the majority of the texts in the Academic Corpus were written between 1993 and 1996, the texts from the LOB and Brown corpora were written in 1961.
2. Obtain more in-depth information about academic vocabulary. Does each of the words in the AWL have roughly the same meaning over a range of subject areas? If not, how can teachers effectively teach learners to recognize distinctions of meaning in different subject areas? Do some lexical items take on a grammatical-type function in texts?
3. Investigate whether learners would be well served by further lists of subtechnical and technical vocabulary in subject areas or whether this knowledge is more easily developed through reading.
4. Investigate the AWL in regard to spoken academic English. Does the AWL, which is based on written academic English, account for spoken academic English, or is this a completely separate genre that needs its own academic word list?

Good knowledge of academic vocabulary is essential for success at higher levels of education (Corson, 1997). By highlighting the words that university students will meet in a wide range of academic texts, the AWL provides the foundation for a systematic approach to academic vocabulary development and may serve as a useful basis for further research into the nature of academic vocabulary.

ACKNOWLEDGMENTS

This research is based on a master's thesis completed at Victoria University of Wellington. I would like to thank my supervisor, Paul Nation, for his helpful comments, suggestions, and encouragement during the writing of the thesis and this article. I am grateful also to Graeme Kennedy for his comments and to *TESOL Quarterly's* reviewers and editor.

THE AUTHOR

Averil Coxhead teaches English for academic purposes in the School of Linguistics and Applied Language Studies at Victoria University of Wellington.

REFERENCES

- Atkins, S., Clear, J., & Ostler, N. (1992). Corpus design criteria. *Literary and Linguistic Computing*, 7, 1–16.
- Bauer, L. (1993). *Manual of information to accompany the Wellington Corpus of Written New Zealand English*. Wellington, New Zealand: Victoria University of Wellington.
- Bauer, L., & Nation, I. S. P. (1993). Word families. *International Journal of Lexicography*, 6, 253–279.
- Biber, D. (1989). A typology of English texts. *Linguistics*, 27, 3–43.
- Biber, D. (1993). Representativeness in corpus design. *Literary and Linguistic Computing*, 8, 243–257.
- Biber, D., Conrad, S., & Reppen, R. (1994). Corpus-based issues in applied linguistics. *Applied Linguistics*, 15, 169–189.
- Biber, D., Conrad, S., & Reppen, R. (1998). *Corpus linguistics: Investigating language structure and use*. Cambridge: Cambridge University Press.
- Buckle, R., Kim, K., & Hall, V. B. (1994). *Dating New Zealand business cycles* (GSBGM Working Paper 6). Wellington, New Zealand: Victoria University of Wellington.
- Campion, M., & Elley, W. (1971). *An academic vocabulary list*. Wellington: New Zealand Council for Educational Research.
- Cohen, A., Glasman, H., Rosenbaum-Cohen, P. R., Ferrara, J., & Fine, J. (1988). Reading English for specialised purposes: Discourse analysis and the use of standard informants. In P. Carrell, J. Devine, & D. Eskey (Eds.), *Interactive approaches to second language reading* (pp. 152–167). Cambridge: Cambridge University Press.
- Collins COBUILD dictionary* (2nd ed.). (1995). London: HarperCollins.
- Corson, D. (1997). The learning and use of academic English words. *Language Learning*, 47, 671–718.
- Coxhead, A. J. (1998). *An academic word list* (English Language Institute Occasional Publication No. 18). Wellington, New Zealand: Victoria University of Wellington.

- Daugherty, C. (1997). *Genetics: A human perspective*. Unpublished manuscript.
- Ellis, R. (1990). *Instructed second language acquisition*. Oxford: Basil Blackwell.
- Engels, L. K. (1968). The fallacy of word counts. *International Review of Applied Linguistics*, 6, 213–231.
- Farid, A. (1985). *A vocabulary workbook: Prefixes, roots, and suffixes for ESL students*. Englewood Cliffs, NJ: Prentice Hall.
- Francis, W. N., & Kucera, H. (1982). *Frequency analysis of English usage*. Boston: Houghton Mifflin.
- Ghadessy, P. (1979). Frequency counts, words lists, and materials preparation: A new approach. *English Teaching Forum*, 17, 24–27.
- Heatley, A., & Nation, P. (1996). Range [Computer software]. Wellington, New Zealand: Victoria University of Wellington. (Available from <http://www.vuw.ac.nz/lals>)
- Hirsh, D. (1993). *The vocabulary demands and vocabulary learning opportunities in short novels*. Unpublished master's thesis, Victoria University of Wellington, New Zealand.
- Hwang, K. (1989). *Reading newspapers for the improvement of vocabulary and reading skills*. Unpublished master's thesis, Victoria University of Wellington, New Zealand.
- Johansson, S. (1978). *Manual of information to accompany the Lancaster-Oslo/Bergen Corpus of British English, for use with digital computers*. Oslo, Norway: University of Oslo, Department of English.
- Kennedy, G. (1998). *An introduction to corpus linguistics*. London: Addison Wesley Longman.
- Kidman, J. (1995). *Dialogues with Maori students: Some implications for academic development* (Higher Education in New Zealand Occasional Paper 2). Wellington: Syndicate of Educational Development Centres of New Zealand Universities.
- Long, M. (1988). Instructed interlanguage development. In L. Beebe (Ed.), *Issues in second language acquisition* (pp. 335–373). New York: Newbury House.
- Lynn, R. W. (1973). Preparing word lists: a suggested method. *RELC Journal*, 4(1), 25–32.
- Miller, J. (1997). Actions for damages for personal injury in New Zealand. In D. Rennie & J. Miller (Eds.), *Brookers' accident compensation in New Zealand* (Suppl.). Wellington, New Zealand: Brookers.
- Murison-Bowie, S. (1993). *MicroConcord manual*. Oxford: Oxford University Press.
- Nagy, W., Anderson, R., Schommer, M., Scott, J. A., & Stallman, A. (1989). Morphological families in the internal lexicon. *Reading Research Quarterly*, 24, 262–281.
- Nation, I. S. P. (1983). Testing and learning vocabulary. *Guidelines*, 5(1), 12–25.
- Nation, I. S. P. (in press). *Learning vocabulary in another language*. Cambridge: Cambridge University Press.
- Praninskas, J. (1972). *American university word list*. London: Longman.
- Richards, J. (1974). Word lists: problems and prospects. *RELC Journal*, 5(2), 69–84.
- Sinclair, J. (1991). *Corpus, concordance and collocation*. Oxford: Oxford University Press.
- Sutarsyah, C., Nation, P., & Kennedy, G. (1994). How useful is EAP vocabulary for ESP? *RELC Journal* 25(2), 34–50.
- Thorndike, E. & Lorge, I. (1944). *The teacher's word book of 30,000 words*. New York: Teachers College Press.
- Valcourt, G., & Wells, L. (1999). *Mastery: A University Word List reader*. Ann Arbor: The University of Michigan Press.
- West, M. (1953). *A general service list of English words*. London: Longman, Green.

- Worthington, D., & Nation, P. (1996). Using texts to sequence the introduction of new vocabulary in an EAP course. *RELC Journal*, 27(2), 1–11.
- Xue, G., & Nation, I. S. P. (1984). A university word list. *Language Learning and Communication*, 3, 215–229.
- Yorkey, R. (1981). *Checklists for vocabulary study*. New York: Longman.

APPENDIX A

Headwords² of the Word Families in the Academic Word List

Numbers indicate the sublist of the Academic Word List (e.g., *abandon* and its family members are in Sublist 8). Sublist 1 contains the most frequent words in the list, and Sublist 10 contains the least frequent.

| | | | | | |
|--------------|----|--------------|----|---------------|----|
| abandon | 8 | aspect | 2 | coincide | 9 |
| abstract | 6 | assemble | 10 | collapse | 10 |
| academy | 5 | assess | 1 | colleague | 10 |
| access | 4 | assign | 6 | commence | 9 |
| accommodate | 9 | assist | 2 | comment | 3 |
| accompany | 8 | assume | 1 | commission | 2 |
| accumulate | 8 | assure | 9 | commit | 4 |
| accurate | 6 | attach | 6 | commodity | 8 |
| achieve | 2 | attain | 9 | communicate | 4 |
| acknowledge | 6 | attitude | 4 | community | 2 |
| acquire | 2 | attribute | 4 | compatible | 9 |
| adapt | 7 | author | 6 | compensate | 3 |
| adequate | 4 | authority | 1 | compile | 10 |
| adjacent | 10 | automate | 8 | complement | 8 |
| adjust | 5 | available | 1 | complex | 2 |
| administrate | 2 | aware | 5 | component | 3 |
| adult | 7 | behalf | 9 | compound | 5 |
| advocate | 7 | benefit | 1 | comprehensive | 7 |
| affect | 2 | bias | 8 | comprise | 7 |
| aggregate | 6 | bond | 6 | compute | 2 |
| aid | 7 | brief | 6 | conceive | 10 |
| albeit | 10 | bulk | 9 | concentrate | 4 |
| allocate | 6 | capable | 6 | concept | 1 |
| alter | 5 | capacity | 5 | conclude | 2 |
| alternative | 3 | category | 2 | concurrent | 9 |
| ambiguous | 8 | cease | 9 | conduct | 2 |
| amend | 5 | challenge | 5 | confer | 4 |
| analogy | 9 | channel | 7 | confine | 9 |
| analyse | 1 | chapter | 2 | confirm | 7 |
| annual | 4 | chart | 8 | conflict | 5 |
| anticipate | 9 | chemical | 7 | conform | 8 |
| apparent | 4 | circumstance | 3 | consent | 3 |
| append | 8 | cite | 6 | consequent | 2 |
| appreciate | 8 | civil | 4 | considerable | 3 |
| approach | 1 | clarify | 8 | consist | 1 |
| appropriate | 2 | classic | 7 | constant | 3 |
| approximate | 4 | clause | 5 | constitute | 1 |
| arbitrary | 8 | code | 4 | constrain | 3 |
| area | 1 | coherent | 9 | construct | 2 |

² Headwords are stem noun or verb forms.

| | | | | | |
|---------------|----|-------------|----|----------------|----|
| consult | 5 | document | 3 | flexible | 6 |
| consume | 2 | domain | 6 | fluctuate | 8 |
| contact | 5 | domestic | 4 | focus | 2 |
| contemporary | 8 | dominate | 3 | format | 9 |
| context | 1 | draft | 5 | formula | 1 |
| contract | 1 | drama | 8 | forthcoming | 10 |
| contradict | 8 | duration | 9 | foundation | 7 |
| contrary | 7 | dynamic | 7 | found | 9 |
| contrast | 4 | economy | 1 | framework | 3 |
| contribute | 3 | edit | 6 | function | 1 |
| controversy | 9 | element | 2 | fund | 3 |
| convene | 3 | eliminate | 7 | fundamental | 5 |
| converse | 9 | emerge | 4 | furthermore | 6 |
| convert | 7 | emphasis | 3 | gender | 6 |
| convince | 10 | empirical | 7 | generate | 5 |
| cooperate | 6 | enable | 5 | generation | 5 |
| coordinate | 3 | encounter | 10 | globe | 7 |
| core | 3 | energy | 5 | goal | 4 |
| corporate | 3 | enforce | 5 | grade | 7 |
| correspond | 3 | enhance | 6 | grant | 4 |
| couple | 7 | enormous | 10 | guarantee | 7 |
| create | 1 | ensure | 3 | guideline | 8 |
| credit | 2 | entity | 5 | hence | 4 |
| criteria | 3 | environment | 1 | hierarchy | 7 |
| crucial | 8 | equate | 2 | highlight | 8 |
| culture | 2 | equip | 7 | hypothesis | 4 |
| currency | 8 | equivalent | 5 | identical | 7 |
| cycle | 4 | erode | 9 | identify | 1 |
| data | 1 | error | 4 | ideology | 7 |
| debate | 4 | establish | 1 | ignorance | 6 |
| decade | 7 | estate | 6 | illustrate | 3 |
| decline | 5 | estimate | 1 | image | 5 |
| deduce | 3 | ethic | 9 | immigrate | 3 |
| define | 1 | ethnic | 4 | impact | 2 |
| definite | 7 | evaluate | 2 | implement | 4 |
| demonstrate | 3 | eventual | 8 | implicate | 4 |
| denote | 8 | evident | 1 | implicit | 8 |
| deny | 7 | evolve | 5 | imply | 3 |
| depress | 10 | exceed | 6 | impose | 4 |
| derive | 1 | exclude | 3 | incentive | 6 |
| design | 2 | exhibit | 8 | incidence | 6 |
| despite | 4 | expand | 5 | incline | 10 |
| detect | 8 | expert | 6 | income | 1 |
| deviate | 8 | explicit | 6 | incorporate | 6 |
| device | 9 | exploit | 8 | index | 6 |
| devote | 9 | export | 1 | indicate | 1 |
| differentiate | 7 | expose | 5 | individual | 1 |
| dimension | 4 | external | 5 | induce | 8 |
| diminish | 9 | extract | 7 | inevitable | 8 |
| discrete | 5 | facilitate | 5 | infer | 7 |
| discriminate | 6 | factor | 1 | infrastructure | 8 |
| displace | 8 | feature | 2 | inherent | 9 |
| display | 6 | federal | 6 | inhibit | 6 |
| dispose | 7 | fee | 6 | initial | 3 |
| distinct | 2 | file | 7 | initiate | 6 |
| distort | 9 | final | 2 | injure | 2 |
| distribute | 1 | finance | 1 | innovate | 7 |
| diverse | 6 | finite | 7 | input | 6 |

| | | | | | |
|--------------|----|-----------------|----|--------------|----|
| insert | 7 | minimise | 8 | precede | 6 |
| insight | 9 | minimum | 6 | precise | 5 |
| inspect | 8 | ministry | 6 | predict | 4 |
| instance | 3 | minor | 3 | predominant | 8 |
| institute | 2 | mode | 7 | preliminary | 9 |
| instruct | 6 | modify | 5 | presume | 6 |
| integral | 9 | monitor | 5 | previous | 2 |
| integrate | 4 | motive | 6 | primary | 2 |
| integrity | 10 | mutual | 9 | prime | 5 |
| intelligence | 6 | negate | 3 | principal | 4 |
| intense | 8 | network | 5 | principle | 1 |
| interact | 3 | neutral | 6 | prior | 4 |
| intermediate | 9 | nevertheless | 6 | priority | 7 |
| internal | 4 | nonetheless | 10 | proceed | 1 |
| interpret | 1 | norm | 9 | process | 1 |
| interval | 6 | normal | 2 | professional | 4 |
| intervene | 7 | notion | 5 | prohibit | 7 |
| intrinsic | 10 | notwithstanding | 10 | project | 4 |
| invest | 2 | nuclear | 8 | promote | 4 |
| investigate | 4 | objective | 5 | proportion | 3 |
| invoke | 10 | obtain | 2 | prospect | 8 |
| involve | 1 | obvious | 4 | protocol | 9 |
| isolate | 7 | occupy | 4 | psychology | 5 |
| issue | 1 | occur | 1 | publication | 7 |
| item | 2 | odd | 10 | publish | 3 |
| job | 4 | offset | 8 | purchase | 2 |
| journal | 2 | ongoing | 10 | pursue | 5 |
| justify | 3 | option | 4 | qualitative | 9 |
| label | 4 | orient | 5 | quote | 7 |
| labour | 1 | outcome | 3 | radical | 8 |
| layer | 3 | output | 4 | random | 8 |
| lecture | 6 | overall | 4 | range | 2 |
| legal | 1 | overlap | 9 | ratio | 5 |
| legislate | 1 | overseas | 6 | rational | 6 |
| levy | 10 | panel | 10 | react | 3 |
| liberal | 5 | paradigm | 7 | recover | 6 |
| licence | 5 | paragraph | 8 | refine | 9 |
| likewise | 10 | parallel | 4 | regime | 4 |
| link | 3 | parameter | 4 | region | 2 |
| locate | 3 | participate | 2 | register | 3 |
| logic | 5 | partner | 3 | regulate | 2 |
| maintain | 2 | passive | 9 | reinforce | 8 |
| major | 1 | perceive | 2 | reject | 5 |
| manipulate | 8 | percent | 1 | relax | 9 |
| manual | 9 | period | 1 | release | 7 |
| margin | 5 | persist | 10 | relevant | 2 |
| mature | 9 | perspective | 5 | reluctance | 10 |
| maximise | 3 | phase | 4 | rely | 3 |
| mechanism | 4 | phenomenon | 7 | remove | 3 |
| media | 7 | philosophy | 3 | require | 1 |
| mediate | 9 | physical | 3 | research | 1 |
| medical | 5 | plus | 8 | reside | 2 |
| medium | 9 | policy | 1 | resolve | 4 |
| mental | 5 | portion | 9 | resource | 2 |
| method | 1 | pose | 10 | respond | 1 |
| migrate | 6 | positive | 2 | restore | 8 |
| military | 9 | potential | 2 | restrain | 9 |
| minimal | 9 | practitioner | 8 | restrict | 2 |

| | | | | | |
|-------------|----|-----------------|----|------------|----|
| retain | 4 | status | 4 | thesis | 7 |
| reveal | 6 | straightforward | 10 | topic | 7 |
| revenue | 5 | strategy | 2 | trace | 6 |
| reverse | 7 | stress | 4 | tradition | 2 |
| revise | 8 | structure | 1 | transfer | 2 |
| revolution | 9 | style | 5 | transform | 6 |
| rigid | 9 | submit | 7 | transit | 5 |
| role | 1 | subordinate | 9 | transmit | 7 |
| route | 9 | subsequent | 4 | transport | 6 |
| scenario | 9 | subsidy | 6 | trend | 5 |
| schedule | 8 | substitute | 5 | trigger | 9 |
| scheme | 3 | successor | 7 | ultimate | 7 |
| scope | 6 | sufficient | 3 | undergo | 10 |
| section | 1 | sum | 4 | underlie | 6 |
| sector | 1 | summary | 4 | undertake | 4 |
| secure | 2 | supplement | 9 | uniform | 8 |
| seek | 2 | survey | 2 | unify | 9 |
| select | 2 | survive | 7 | unique | 7 |
| sequence | 3 | suspend | 9 | utilise | 6 |
| series | 4 | sustain | 5 | valid | 3 |
| sex | 3 | symbol | 5 | vary | 1 |
| shift | 3 | tape | 6 | vehicle | 8 |
| significant | 1 | target | 5 | version | 5 |
| similar | 1 | task | 3 | via | 8 |
| simulate | 7 | team | 9 | violate | 9 |
| site | 2 | technical | 3 | virtual | 8 |
| so-called | 10 | technique | 3 | visible | 7 |
| sole | 7 | technology | 3 | vision | 9 |
| somewhat | 7 | temporary | 9 | visual | 8 |
| source | 1 | tense | 8 | volume | 3 |
| specific | 1 | terminate | 8 | voluntary | 7 |
| specify | 3 | text | 2 | welfare | 5 |
| sphere | 9 | theme | 8 | whereas | 5 |
| stable | 5 | theory | 1 | whereby | 10 |
| statistic | 4 | thereby | 8 | widespread | 8 |

APPENDIX B

Sample Texts From the Academic Corpus³

Text From the Commerce Subcorpus (Buckle, Kim, & Hall, 1994)

DATING NEW ZEALAND BUSINESS CYCLES

I. Introduction

Dating the turning points and duration of business cycles has long been associated with the construction of aggregate reference cycle indexes, and their associated leading, coincident and lagging indicators. This was along lines originally developed by Burns and Mitchell (1946), and subsequently by colleagues at the National Bureau of Economic Research (NBER), e.g. Klein (1990). More recently, identifying the turning points and duration of business cycles has been an important aspect of two further areas of business cycle research: the evaluation of theoretical and associated empirical business cycle models, e.g. King and Plosser (1994), Simkins (1994); and the analysis of the time varying characteristics of business cycles, e.g. Diebold and Rudebusch (1992), Watson (1994).

³ Words in the Academic Word List are underlined.

The Burns and Mitchell technique of dating business cycles relied primarily on two sorts of information: the descriptive evidence from business publications and general business conditions indices, and the “specific cycles” found in many individual series and the tendency for turning points to sometimes cluster at certain dates. Based on this information, a set of reference cycle dates were selected that specified the turning points in “aggregate economic activity”. A key feature of the Burns and Mitchell approach was to focus on the amount of cyclical co-movement or coherence among a large number of economic variables. This co-movement is the prime characteristic of their definition of the business cycle: “. . . a cycle consists of expansions occurring at about the same time in many economic activities, followed by similarly general recessions, contractions, and revivals which merge into the expansion phase of the next cycle; . . . in duration business cycles vary from more than one year to ten or twelve years . . .” (Burns and Mitchell, 1946, p 3).

The NBER approach is based on the view that there is no unique way of combining all these activities, and accordingly the business cycle cannot be fully depicted by a single measure, e.g. Burns (1969, p 13). Burns and Mitchell, and subsequent NBER researchers, intended therefore, before the computer age, to provide a standard technique with a set of decision rules for deriving business cycle turning points based on these two sorts of information. In practice, this involved the application of a standard format of filtering procedures to extract the turning points in each data series, and then combining this information in a judgemental way to determine a single turning point date. Other procedures, notably reference cycle indexes and coincident indexes, subsequently emerged as supplementary procedures for combining a large number of data series including various measures of output, production inputs, price series, monetary aggregates, etc, into a single composite index which have also been used to identify turning points.

Text From the Science Subcorpus (Daugherty, 1997)

Transmission Genetics

Gregor Mendel’s experiments, described in Chapter 4, are models of scientific elegance. Mendel reported his studies on the inheritance of seven different characteristics of pea plants by following the transmission of these traits from parent to offspring. Although he knew nothing of the chemical nature of genes, Mendel was able to describe aspects of their function quite accurately. These principles became the starting point of the new science. The study of transmission of single genes and traits is sometimes called Mendelian genetics. The following six principles reflect contemporary concepts in transmission genetics.

1. Inheritance is the transmission of traits and characteristics from one generation to the next. In his experiments, Mendel showed that some simple traits are governed by the effects of two factors, one inherited from each parent. We now call these factors genes.
2. Differing chemical forms of a gene that govern a single function or trait are called alleles. One allele of each gene is carried by each egg and each sperm that unite to form a new individual. Thus, a newly fertilized egg, or zygote, will contain a complete set of paired alleles that governs the development of the traits of the new individual. Sometimes, the words *gene* and *allele* are used interchangeably. Individuals who have two identical alleles of a gene are said to be homozygous. Individuals who have different alleles of a gene are heterozygous.
3. All of the genes (or alleles) that an individual has are called the *genotype* or the *genome* of that person. Genotype may also refer to the pair of alleles, one derived from each parent, that govern a single trait.
4. Characteristics of an individual, singly or collectively, are called the *phenotype* of the person (figure 1.6a). The phenotype results from genes being expressed in a particular environment. Distinguishing the effects of the genes from the effects of the environment in determining the phenotype has proven to be extremely difficult.
5. The two alleles that govern a single trait may interact. Sometimes, one allele is expressed, and the other is not. In such a case, the allele that is expressed is said to be dominant to the recessive allele which is not expressed. Alleles that are both expressed are said to be co-dominant.

6. Mendelian rules govern inheritance not only in pea plants, but also humans, other animals, and plants. Mendelian mechanisms of inheritance are shared by virtually all higher organisms. Simpler, single-celled organisms such as bacteria do not follow Mendelian patterns of inheritance. Bacterial genes occur singly rather than in pairs.

Cytogenetics

Mendel based his ideas on experimental results using whole plants. Quite independently, biologists analyzed the structure of cells and their components microscopically. By 1903, the basic stages of cell division had also been identified, and genes were known to occur on chromosomes. The exciting result was that the behavior of chromosomes found in the nucleus of a cell could explain many of Mendel's findings. Cytogenetics is the study of the structure and function of chromosomes, and of cells during division.

7. Genes occupy segments of *chromosomes*, dark-staining bodies that occur in the nucleus of cells. Chromosomes (figure 1.7) are usually visible only during, or immediately before, cell division. The precise site that a gene occupies on a chromosome is called a genetic locus. In Mendelian terms, a genetic locus on a chromosome is occupied by an allele.

Text From the Arts Subcorpus (Kidman, 1995)

OVERVIEW

The American educator Maxine Greene (1984) has written of the relationship between students and teachers:

There is a danger in the tendency to disconfirm their experiences and responses, because they do not participate in what we believe to be "literate" discourse and because they often do not value what we take for granted to be valuable. (p. 293)

Some students believed that the discourses which allow academics within a discipline to speak to one another, or in communicating primarily through 'professional' categories, the life stories or identities of the speakers can become hidden. Some believed that the construction of academic discourses provides teachers with elaborate languages which allow a degree of intellectual and personal concealment. One student said, "I get lost in the jargon, we don't converse here at the University or share our common understandings. We swop abstract principles and call it an education". Maori students who have come to the University seeking new understandings or who wish to combine their own cultural knowledge with academic meanings sometimes find that the patterns of academic discourse inhibits them from finding the words for their own lived experiences. The 'human' face of learning, described by another student seems to disappear in the sea of faces, making communication uneasy or common values difficult to identify.

Part of this distancing effect also lies in the physical layout of the University and its classrooms. The immovable rows of seats in the larger lecture theatres are not conducive to the development of interactive learning environments. Some lecturers continue to give classes without the aid of media in the belief that their ancient lecture notes and measured tones of authority will motivate 150 or 300 recently enrolled First Years to investigate a subject further at a later date. In spite of this however, several students had attended lectures where staff had altered this traditional approach to large group teaching and, despite the problems with room layout and class size, had provided interactive instruction which the students remembered and enjoyed.

Small group or tutorial situations were a greatly preferred style of teaching. Several students mentioned that they initially lacked confidence in speaking before people they didn't know, but they had gained a sense of self assurance over their time at the University and most of them used tutorials to "bounce ideas off other students". Furthermore, some students had extended the study group approach independently and had formed informal, self-led study groups with other Maori students. These groups hold particular significance in contributing to the informal rhythms of university life. They facilitate the development of strong networks among Maori students, and they also provide support for students who have knowledge of their own culture and who wish to enter the 'deep' structures of learning. For those students competing against the pressures of university study, the commitment to adopt 'deep' approaches to their learning can be swayed by external factors, such as departmental ethos, teaching methodologies and an

overabundance of unconnected information. The desire to create knowledge from a quantity of information may be in part satisfied by students who participate in self-directed study groups. Here, it may be possible for the university to lend its support to those students who are taking the time to extract an understanding of intellectual processes alongside the demand to keep producing essays and degrees.

Text From Law Subcorpus (Miller, 1997)

ACTIONS FOR DAMAGES FOR PERSONAL INJURY IN NEW ZEALAND

Introduction

This chapter is published as a supplement to Brooker's "Accident Compensation in New Zealand". It is intended to direct the attention of persons advising accident victims to the need to explore the possibility of bringing common law actions for damages in appropriate circumstances. Clearly the topic cannot be covered comprehensively in a chapter of this kind but attention is drawn to the more obvious cases where it would appear common law actions might be available in New Zealand Courts. Readers may also wish to refer to the following recent material, R Harrison, Matters of Life and Death, Legal Research Foundation No 35, Auckland, 1993; S Todd and J Black, Accident Compensation and the Barring of Actions for Damages (1993) 2 Tort Law Review 197

1. Statutory Compensation or damages?

The lawyers duty

1.1 Because of the limited compensation available under the ARCI Act, lawyers and others who advise persons on their rights arising out of personal injury, should not limit their efforts to attempting to prove that the injury is covered by the Act. They have a duty to look for ways of showing that the ARCI Act does not apply to the injuries.

1.2 In order to ensure that the client receives the maximum benefits available, it is the lawyer's duty when advising a person who has suffered personal injury, to explore the possibility of an action for common law damages. Neglecting to explore this option may amount to professional negligence on the part of a barrister or solicitor. (see Keys v. St. L Reeves A55/85 H.C. New Plymouth 13th April 1992, Smellie J.)

Limited prohibition against suing

1.3 Apart from the actions for loss of consortium and loss of services, the Accident Compensation legislation has never removed the common law right of action for damages for personal injury in New Zealand. Section 5 of the 1972, and s27 of the 1982 Accident Compensation Acts, prevented the bringing of proceedings for damages in a New Zealand Court, but there was no prohibition against common law proceedings in a Court outside New Zealand in respect of personal injury suffered in New Zealand. Nor, apart from the two actions mentioned beforehand, did the legislation remove the common law right where a cause of action arose in New Zealand. It imposed a procedural bar to suing in a New Zealand Court to recover compensatory damages. It was however, possible to sue for "exemplary" or punitive damages (see Donselaar v Donselaar [1982] 1 NZLR 87 and Auckland City Council v Blundell [1986] 1 NZLR 732(CA)).

1.4 In common with the previous legislation, s14 of the ARCI Act 1992, prohibits the bringing in any New Zealand Court of proceedings for damages arising directly or indirectly out of personal injury covered by the Act or personal injury by accident covered by the 1972 or 1982 Accident Compensation Acts. Like its predecessors the prohibition in s 14 does not apply to exemplary or punitive damages.