

ESOL Online

NCEA Vocabulary

The Development of an NCEA Corpus and Word List

The National Certificate of Educational Achievement (NCEA) has been progressively introduced into New Zealand secondary schools since 2001. The NCEA corpus was developed during 2002 and 2003 by David Wallace to assist senior New Zealand secondary school students in their comprehension of the tasks involved in the NCEA assessment regime. This introduction gives a brief outline of how the NCEA Corpus was constructed and the subsequent development of the word list. More detail can be found in the forthcoming volume of TESOLANZ.

The words for the NCEA Word List were extracted from a corpus of level one and two NCEA internal and external assessments downloaded from two sites. [Te Kete Ipurangi](#) and [NZQA](#). Once the teacher information, marking schedules, generic NZQA information and any foreign words were removed the corpus came to a total of over 500,000 words. For analysis the NCEA Corpus was divided into four sub-corpora representing the major study areas of the secondary school curriculum.

The range and number of NCEA assessments in the NCEA Corpus

Discipline	Number of subjects	Number of internal assessments	Number of external assessments	Total assessments	Total number of words
Arts	20	172	98	270	243,724
Sciences	8	173	114	287	165,754
Fine Arts and Technology	6	99	44	143	91,023
Commerce	2	28	29	57	37,513
TOTAL	37	472	265	757	538,014

The words for the NCEA word list were chosen using the following three criteria:

- The word must be outside the [first 2000 words](#).
- Range - Members of the word family had to occur in at least three of the sub-corpora. This was to avoid specialised technical words of just one subject or area.
- Frequency - Members of the word family had to occur at least 25 times in the corpus.

This resulted in a word list of 264 word families. These were then divided into six sub-lists ranging in frequency from sub-list one (most frequent) to sub-list six (least frequent).

The NCEA word list is approximately half the length of the [Academic Word List](#) (Coxhead, 2000), yet the words contained in it still account for more than 11% of the total words in the corpus.

The Frequency of Words in the NCEA Corpus



Some of the word families found in word lists like the [Academic Word List](#) (Coxhead, 2000) are large and a student could spend a considerable amount of time learning words that they are unlikely to meet outside of the language classroom. To address this issue, frequency and range information were used to limit the word forms to the most common and widely used. During the process some words from the word list were removed. *Merit*, for instance, was found only to refer to the achievement criteria. For example, "to gain a merit grade in this assessment you must complete all the questions." Since a student completing NCEA would be likely to be already aware of this use it was removed from the list. Some morphological variants were also removed from the word families because they did not occur in the corpus. These included highly infrequent terms, at least in secondary school English, such as *conceptualisation and uncontextualised*. Words with prefix stems in the word family were only retained if they actually occurred in the corpus. Also, verb conjugations and noun plurals were removed. It was assumed that a student studying from the NCEA wordlist would know how to conjugate a regular verb, make a plural, and would understand the meaning of a common prefix. An example of a simplified entry is as follows:

First NCEA Word List

interpret

interpretation

interpretations

interpretative

interpreted

interpreting

interpretive

interprets

misinterpret

misinterpretation

misinterpretations

New NCEA Word List

interpret

interpretation

interpretative

misinterpreted
 misinterpreting
 misinterprets
 reinterpret
 reinterpreted
 reinterprets
 reinterpreting
 reinterpretation
 reinterpretations

Each member of the word family was then analysed for frequency and given a star rating based upon the number of occurrences in the corpus.

The star frequency rating system used in the NCEA word list

FREQUENCY IN THE NCEA CORPUS:	Star Rating	Frequency
	*****	200 or more occurrences
	****	100 – 199
	***	50 - 99
	**	20 - 49
	*	1 - 20
	(no stars)	does not occur

The final entry for the word family of *interpret* appears as follows:

The NCEA wordlist entry for the word family - *interpret*

interpret	v	***
interpretation	n	***
interpretative	adj.	*

The NCEA word list is primarily designed for students to self-select words for intensive vocabulary study both inside and outside of the ESOL classroom. I understand concerns about washback and assessment led teaching, however I have found the NCEA list to be a valuable tool to focus student learning on vocabulary that will be of the most use to them. I hope that by its wider dissemination that more teachers and students will benefit from its use.

David Wallace
HOD ESOL
Hutt Valley High School
d.wallace@xtra.co.nz

November 2003

Coxhead, Averill. 2000. "A New Academic Word List." *TESOL Quarterly* 34,2. 213-238.

Nation, I.S.P. 2001. *Learning Vocabulary in Another Language*. Cambridge: Cambridge University Press.

© Ministry of Education, Wellington, New Zealand

Printed on 07 Jun 2009.