WHICH WORDS MUST STUDENTS KNOW?:
A NOTE ON "A NEW ACADEMIC WORD LIST" (Coxhead 2000)

Coxhead (2000) reports on a new **Academic Word List** (AWL) which is of great interest for students and their teachers.

The words in AWL are ones with which university students of English must be thoroughly familiar for both reading and writing academic prose.

AWL contains 570 "words" which have both high frequency and high coverage in academic texts, independent of subject area and discipline. These 570 "words" consist of head-words plus their inflected and derived forms, and therefore around 3,100 word-forms altogether, e.g.:

- **concept**: conception, concepts, conceptual, conceptualisation, conceptualise, conceptualised, conceptualises, conceptualising, conceptually.

(It does not contain both -*ise* and -*ize* spellings. This is largely irrelevant for teaching, though it could be important if the list is used for computer text-processing.)

To prepare the list, Coxhead used a corpus of 3.5 million words, plus software which can calculate how often a word occurs (its frequency), and how many different texts in the corpus it occurs in (its range). The texts are from academic journals and university textbooks from four main areas: arts, commerce, law, and natural science.

AWL does not include any of the 2,000 most frequent words in English, listed in the General Service List (West 1953), but only words outside these very frequent words. To be included in AWL, a word had to occur:

- at least 100 times altogether in the whole academic corpus
- at least ten times in each of the four sub-corpora
- in at least half of 28 more finely defined subject areas, e.g. history, linguistics, economics, management, biology, geography.

AWL gives very good coverage of academic texts, independent of subject area. Here it must be remembered that words are very uneven in their frequency. In a typical academic text

- the single word *the* covers around 6 or 7% of running text
- the top ten words cover over 20%
- the 2,000 most frequent words cover around 75%.

The words in AWL typically cover a further 10%. The remaining 15% will be specialized words which are specific to a given subject area or topic, plus proper names, etc. (Conversely, these are words which are not frequent in fiction: they cover only around 1.4% of fictional texts.)

FULL DETAILS OF AWL

The 570 head-words are published in Coxhead (2000). The head-words, alone and with their associated words, grouped in various ways – plus a lot of supplementary information, publications, etc, – are on Coxhead's web-site:

http://language.massey.ac.nz/staff/awl/

The complete word-list (head-words plus all derived and inflected forms: around 3,100 word-forms) is also available as part of a zipped package at Paul Nation's website:

http://www.victoria.ac.nz/lals/staff/paul-nation/nation.aspx

This package also contains associated software, written by Paul Nation and his colleagues, which analyses word frequency and range, and compares texts and corpora with each other and with word frequency lists. (Nation 2001.)

HEAD-WORDS

AWL is divided into ten sub-groups, from most to least frequent. This is the form of the lists below. I have taken the 570 head-words (without derived forms, etc), and listed them separately in ten groups from 1 (most frequent) to 10 (least frequent – but still essential for reading academic texts).

Here again, coverage is very uneven. Group 1 covers 3.6% of the corpus.

This means that if you are reading academic prose, you could expect to come across each word in group 1, on average, once every four pages or so.

Coverage then falls off rapidly. Groups 1 to 5 inclusive cover over 8% of the corpus. Group 9 covers only 0.2%, and group 10 only 0.1%.

AWL is, of course, just a list, not teaching materials. Even the bare list, however, can provide a check – for students or teachers – on what words students know.

Some descriptive questions (e.g. how consistent are dictionaries in listing derived forms?) and some pedagogical questions (e.g. how are derived forms stored in the mental lexicon?) are discussed by Bauer and Nation (1993).

ACKNOWLEDGEMENTS AND REFERENCES

I am grateful to Averil Coxhead for permission to reproduce the version of AWL below.

Bauer, L & Nation, P (1993) Word families. *International Journal of Lexicography*, 6, 4: 253-79.
Coxhead, A (2000) A new academic word list. *TESOL Quarterly*, 34, 2: 213-238.
Nation, P (2001) Using small corpora to investigate learner needs. In M Ghadessy et al eds *Small Corpus Studies and ELT*. Amsterdam: Benjamins. 31-45.
West, M (1953) *A General Service List of English Words*. London: Longman.

THE ACADEMIC WORD LIST (adapted from Coxhead 2000)

The 570 head-words in AWL the list in Coxhead (2000) are here ordered by frequency, from "group 1" (most frequent) to "group 10" (least frequent).

group 1

analyse approach area assess assume authority available benefit concept consist constitute context contract create data define derive distribute economy environment establish estimate evident export factor finance formula function identify income indicate individual interpret involve issue labour legal legislate major method occur percent period policy principle proceed process require research respond role section sector significant similar source specific structure theory vary

group 2

achieve acquire administrate affect appropriate aspect assist category chapter commission community complex compute conclude conduct consequent construct consume credit culture design distinct element equate evaluate feature final focus impact injure institute invest item journal maintain normal obtain participate perceive positive potential previous primary purchase range region regulate relevant reside resource restrict secure seek select site strategy survey text tradition transfer

group 3

alternative circumstance comment compensate component consent considerable constant constrain contribute convene coordinate core corporate correspond criteria deduce demonstrate document dominate emphasis ensure exclude framework fund illustrate immigrate imply initial instance interact justify layer link locate maximise minor negate outcome partner philosophy physical proportion publish react register rely remove scheme sequence sex shift specify sufficient task technical technique technology valid volume

group 4

access adequate annual apparent approximate attitude attribute civil code commit communicate concentrate confer contrast cycle debate despite dimension domestic emerge error ethnic goal grant hence hypothesis implement implicate impose integrate internal investigate job label mechanism obvious occupy option output overall parallel parameter phase predict principal prior professional project promote regime resolve retain series statistic status stress subsequent sum summary undertake

group 5

academy adjust alter amend aware capacity challenge clause compound conflict consult contact decline discrete draft enable energy enforce entity equivalent evolve expand expose external facilitate fundamental generate generation image liberal licence logic margin medical mental modify monitor network notion objective orient perspective precise prime psychology pursue ratio reject revenue stable style substitute sustain symbol target transit trend version welfare whereas

group 6

abstract accurate acknowledge aggregate allocate assign attach author bond brief capable cite cooperate discriminate display diverse domain edit enhance estate exceed expert explicit federal fee flexible furthermore gender ignorant incentive incidence incorporate index inhibit initiate input instruct intelligence interval lecture migrate minimum ministry motive neutral nevertheless overseas precede presume rational recover reveal scope subsidy tape trace transform transport underlie utilise

group 7

adapt adult advocate aid channel chemical classic comprehensive comprise confirm contrary convert couple decade definite deny differentiate dispose dynamic eliminate empirical equip extract file finite foundation globe grade guarantee hierarchy identical ideology infer innovate insert intervene isolate media mode paradigm phenomenon priority prohibit publication quote release reverse simulate sole somewhat submit successor survive thesis topic transmit ultimate unique visible voluntary

group 8

abandon accompany accumulate ambiguous append appreciate arbitrary automate bias chart clarify commodity complement conform contemporary contradict crucial currency denote detect deviate displace drama eventual exhibit exploit fluctuate guideline highlight implicit induce inevitable infrastructure inspect intense manipulate minimise nuclear offset paragraph plus practitioner predominant prospect radical random reinforce restore revise schedule tense terminate theme thereby uniform vehicle via virtual visual widespread

group 9

accommodate analogy anticipate assure attain behalf bulk cease coherent coincide commence compatible concurrent confine controversy converse device devote diminish distort duration erode ethic format founded inherent insight integral intermediate manual mature mediate medium military minimal mutual norm overlap passive portion preliminary protocol qualitative refine relax restrain revolution rigid route scenario sphere subordinate supplement suspend team temporary trigger unify violate vision

group 10

adjacent albeit assemble collapse colleague compile conceive convince depress encounter enormous forthcoming incline integrity intrinsic invoke levy likewise nonetheless notwithstanding odd ongoing panel persist pose reluctance so-called straightforward undergo whereby

Michael Stubbs
FB2 Anglistik, University of Trier, D-54286 Trier, Germany

Last up-dated February 2008.