**Acquiring a scholar’s voice: Vietnamese students mastering academic vocabulary in thesis writing**

**Sandra Kies PhD** [www.rhetory.com/wrabv](http://www.rhetory.com/wrabv)

**Benedictine University skies@ben.edu**

**Lisle, Illinois, USA.**

**WRAB V (online) 5 March 2021, 1.45-2.00pm**

**Introduction**

Using a scholar’s voice in academic writing at the level of vocabulary involves choosing not only the most accurate technical vocabulary, but also using formal “academic” words and formulaic phrases that are common across many disciplines, while avoiding vocabulary that is considered conversational, colloquial, or intimate (Hyland and Tse, 2007; Leedham and Cai, 2013.)

Various wordlists exist (e.g. West, 1953; Nation, 2000; Coxhead, 2000)

Gardner and Davies (2014) – developed a more refined academic wordlist (NAVL) using the COCAA corpus that provides a more accurate general academic list and also field specific lists.

**The MA Linguistics (TESOL) program**

2-year MA Linguistics (TESOL) program of Benedictine University in partnership with Vietnam National university (University of Social Sciences and Humanities) in Ho Chi Minh City, Vietnam, and delivered in hybrid mode (partially face-to-face in Vietnam and partially online.) Benedictine instructors visit Vietnam every summer to teach intensively.

As part of a 32 credit hr requirement, the program uses an innovative scaffolded thesis model

Students develop their academic writing through four 1 credit thesis writing seminars interwoven through the academic program.

**The MA LING (TESOL) students**

Students are trained English teachers with 2 or more years’ experience.

Employed in schools (K-12), universities or in IELTS or TOEIC preparation centers; Many also “moonlight” as private tutors.

Students value the MA thesis for developing academic writing, and as a strategic advantage in the job market.

Survey showed most of our Grad. students have not written more than 5,000 words before and also believe their vocabulary for academic writing is inadequate.

They also recognize the importance of vocabulary for their own students, judging by the number of thesis topics focused on vocabulary or vocabulary learning issues.

So far, 3 cohorts have graduated (N-43)

Attrition rate is very low; some students drop for a semester or longer. but most return.

Students’ general lang proficiency - IELTS 6.00-8.00 (av. 6.5 = CEFR upper B2- C1); expected to achieve IELTS 6.5 before graduation.

**Scaffolded pedagogy in the MA Ling. (TESOL)**

Students examine and analyze models of academic writing using a genre-based approach (Swales, 1990), practice writing annotated bibliographies, critiques, and other sub-genres of the thesis writing process.

Students receive intensive feedback via D2L (Turnitin/Grademark) at all stages of their thesis writing.

Extensive re-writing takes place.

Students are encouraged to develop academic vocabulary through “noticing” (Ellis, 2006) vocabulary use in target academic genres, and through working with corpora (e.g. COCA word and phrase https://www.wordandphrase.info/wap\_coca.pdf)

**The Research Project**

Aims to inform writing pedagogy in English for Academic Purposes with the goal of developing more targeted instruction in using academic vocabulary in context.

This paper presents mostly quantitative data on a small sub-section of an ongoing mixed methods (Riazi and Candlin, 2014) longitudinal study on academic writing by graduate students writing theses as part of an MA in Linguistics (TESOL), in English, in Vietnam.

**Vocabulary research questions**

What is the vocabulary level of students at the beginning of the thesis writing process compared with reference corpora?

How much change in vocabulary use occurs over the period of drafting? (6mths to one year).

What changes have occurred?

What does this say about our scaffolded thesis writing process with respect to developing academic vocabulary in a discipline?

**Method**

Here I focus on one aspect, the development of academic vocabulary, comparing students’ control over words in the New Academic Vocabulary List(NAVL) (Gardner & Davies, 2014)

So far, analysis has focused only on the draft and final versions of the theses’ introductions discussions and conclusions. These sections are considered most likely to contain students’ most diverse vocabulary, with less reliance on the vocabulary of source texts, or the more formulaic vocabulary found in methods and findings section.

Previously graded texts were anonymized, analyzed by cohort, both individually and collectively using Laurence Anthony’s *Antwordprofiler* andcompared for frequency of lemmas (both core vocabulary and technical vocabulary) listed in the New Academic Vocabulary List (Gardner and Davies, 2014.)

Student files were compared against two reference corpora: The linguistics subcorpus of the BAWE Corpus of student academic Writing, and COCAA 2010-12.

Education and Humanities and Social Sciences lemmas were drawn from the NAVL to make comparisons between the reference corpora and VNL students writing.

**Student writing sample VNL C1 first draft
(Antwordprofiler)**



**Reference corpora** BAWE linguistics subcorpus: level 4 graduate student papers (31 students; 88,433 words) – most comparable to VNL students

NS English L1 (38,222 words)

NNS L1 other than English - living and studying in England- immersed in target language. (50,211 words)

COCAA (academic sub-corpus of corpus of Contemporary American English) 2010-2012 (9,588,174.) Consists of whole papers including reference lists.

Many disciplines; wider range of technical words than VNL texts.

Gives a reference point for target professional writing.

**Comparing draft Intro, Discussion and Conclusions in VNLC1-3**

In the next slide we can see that between first and final drafts VNL students increased overall word count by approx. 50%.

Students collectively gained 115 new core academic words and 41 new discipline-specific words (types = distinct lemmas.)

Percentage of **NAVL core academic words** increased very slightly (13.96-14.06%)

Percentage of **NAVL-discipline-specific words** increased very slightly (17.8-18.03%)

The number of academic words (lemmas) increased, though they went down as a percentage of all words. **NAVL core 1104->1219; NAVL discip. 514->555.** % of “other” words increased (63.86-> 65.08%)



**How did VNL students compare with BAWE?**



**Comparing draft Intro, Discussion and Conclusions in VNL C1-3 (Antwordprofiler)**

**How did VNL students compare with BAWE?**

**BAWE NS (L1 English) compared with VNL students**

NAVL core academic words as % of all text **15.88%** (cf. 13.96->**14.06%**)

NAVL-discipline-specific words as % of all text **14.51%** (17.8->**18.03%**)

Other academic words (lemmas) as a percentage of all text **68.08%** (cf. 63.86-> **65.08**%)

**BAWE NNS (L1 not English) compared with VNL students**

NAVL core academic words as % of all text **15.13%** (cf. 13.96->**14.06%**)

NAVL-discipline-specific words as % of all text **14.81%** (17.8->**18.03%**)

Other academic words (lemmas) as as % of all text **71.02%** (cf. 63.86-> **65.08%**)

**VNL comparable on core ac. words; VNL higher on discipline specific; VNL lower on % of other words.**

**Conclusion**

At a general level, students in all 3 cohorts compare very favorably with NNS in the BAWE corpus (predominantly Mandarin L1 speakers) in the % use of core academic lexis and technical lexis of their field (Hum and Ed, and Soc. Sciences).

There is definitely progression in the overall length of writing and expanded range of words (types= lemmas.)

It is not yet known how much of the expansion comes from more variation in lemmas from a particular word family, e.g., *consider, considerably, consideration, considering, considerable* (general), *reconsideration* (law) or from introducing lemmas from other word families.

**Where to from here?**

Beyond the scope of this preliminary investigation, but we are investigating:

Use of conversational and somewhat aggressive discourse markers such as on *the contrary* or *besides* is common. Comparison with COCAA 2010-12 (9 mill. words) shows *besides* used only 22 times. Our students used *besides* 21 times just in their introductions.

“Overuse” of discourse markers in first position as textual themes. e.g. *moreover, therefore, hence* compared with more professional writers.

Overuse of N-grams such as*…According to, On the one/other hand, On the contrary.*

Unnecessary repetition of words with almost the same meaning in one sentence (lack of understanding of the word meanings. *According to x, he says*…

Overuse of “lexical teddy bears” such as *in general* … *specifically/in particular.*

Use of an appropriate word family, but the wrong derivational suffix.

Appendix 1: New Academic Vocabulary List (NAVL)

 Contains the 20,000 or so most frequent words in 120 million words of the COCA Academic texts (<http://corpus.byu.edu/coca>)

**A. Core.** The top 3,000 "core academic" words (lemmas) in COCA-Academic (e.g. #305 yellow below). To be considered a "core academic word", it must:

1) Occur at least 50% more frequently in the academic portion of COCA than would otherwise be expected (per million words). In other words, 1.50 or higher in the [ratio] column.

2) Have a good "dispersion" [disp column] across the nine domains of academic (a Juilland "d" measure of at least 0.80, for those who know what that means)

3) Have at least 20% of the "expected" frequency in at least seven of the nine domains

4) Not occur more than three times as much as "expected" in any of the nine domains

Appendix 1 contd.

**B. Technical / Domain specific:** About 8,000 other words that occur at least three times as frequently in one of the nine domains of COCA-Academic (e.g. Law or Medicine or Technology) as would otherwise be expected (based on the size of that domain). For example, church occurs more than three times as frequently in [Religion / Philosophy] domain as would otherwise be expected (based on the size of that domain).

**C. General:** About 9,000 other words, which are not "academic" per se (e.g. the, people, new, least), but which are in the top 20,000 lemmas of COCA-Academic (e.g. #307 "least"). The words are further grouped into "bands" of 500 words each.

**References**

Anthony, L. (2020). AntWordProfiler(Version 1.5.0)[Computer Software]. Tokyo, Japan: Waseda University. Available from [http://www.antlab.sci.waseda.ac.jp](http://www.antlab.sci.waseda.ac.jp/)

Coxhead, A. (2000). A new academic wordlist. TESOL Quarterly, 34(2), 213-238. http://dx.doi.org/10.2307/3587951. URL <http://onlinelibrary.wiley.com/doi/> 10.2307/3587951/abstract.

Davies, M. (2008). *The Corpus of Contemporary American English (COCA): 560 million words, 1990-present.* Retrieved from Available online at <https://corpus.byu.edu/coca/>

Durrant, P. (2016). To what extent is the Academic Vocabulary List relevant to university student writing? *English for Specific Purposes (New York, N.Y.)*, *43*, 49–61. <https://doi.org/10.1016/j.esp.2016.01.004>

Ellis, R. (2006) Current Issues in the Teaching of Grammar: An SLA Perspective. *TESOL Quarterly.* 40(1) 83-107..
Gardner, D., & Davies, M. (2014). A new academic vocabulary list. *Applied Linguistics*, 35(3), 305-327.

Hasselgren, A. (1994). Lexical teddy bears and advanced learners: A study into the ways Norwegian students cope with English vocabulary. International Journal of Applied Linguistics, 4(2), 237–258.

Hyland, K., & Tse, P. (2007). Is there an ‘Academic vocabulary’? TESOL Quarterly, 41(2), 235-253. http://dx.doi.org/10.2307/40264352. URL [http://www.jstor](http://www.jstor/). org/stable/40264352.

O’Donnell, M. (2019). *UAM CorpusTool*. Madrid Universidad Autonoma de Madrid. Retrieved from [http://wwwcorpustool.com](http://wwwcorpustool.com/)

Leedham, M., & Cai, G. (2013). Besides … on the other hand: Using a corpus approach to explore the influence of teaching materials on Chinese students’ use of linking adverbials. *Journal of Second Language Writing*, *22*(4), 374–389. <https://doi-org.libweb.ben.edu/10.1016/j.jslw.2013.07.002>

Nation, I. S. P. (2000). Learning vocabulary in another language. Cambridge: Cambridge University Press.

Riazi A.M., & Candlin, C.N. (2014). Mixed method research in language teaching and learning: opportunities, issues and challenges. *Language Learning,* 47(2), 135-173.

Swales, J.M. (1990). *Genre analysis: English in academic and research settings*. Cambridge N.Y. Cambridge University Press.

West, M. (1953). A general service list of English words. London: Longman

The student reference corpus data in this study come from the British Academic Written English (BAWE) corpus, which was developed at the Universities of Warwick, Reading and Oxford Brookes under the directorship of Hilary Nesi and Sheena Gardner (formerly of the Centre for Applied Linguistics, Warwick), Paul Thompson (formerly of the Department of Applied Linguistics, Reading) and Paul Wickens (School of Education, Oxford Brookes), with funding from the ESRC (RES-000-23-0800). [www.coventry.ac.uk/bawe/](http://www.coventry.ac.uk/bawe/)

PPT will also be available on [www.rhetory.com/wrabv](http://www.rhetory.com/wrabv)