

# Lexical interrelatedness of semi-popularization articles across agricultural subdisciplines

Kayo Yamamoto (University of Miyazaki)

Tamao Araki (University of Miyazaki)

Richard S. Lavin (Prefectural University of Kumamoto)

## 1. Background

Regarding the discussion on disciplinary specificity in academic word lists (Hyland & Tse, 2007), discipline-specific word lists have been proposed in various academic fields by researchers (e.g., Martínez, Beck, & Panza, 2009; Ward, 2009; Chen & Ge, 2007; Wang, Liang, & Ge, 2008; Yang, 2015). Word lists for subdisciplines in an academic field have also been proposed, and their interrelationship has been examined. Gilmore and Millar (2018) compiled and examined a civil engineering corpus comprising papers from 11 civil engineering subdisciplines. Using a cluster analysis, they showed lexical correlations among these subdisciplines

A corpus related to a specific area often comprises research articles and is thus challenging to use in the undergraduate curriculum. Muñoz (2015) focused on semi-popularization articles in her research for undergraduate English for Specific Purposes (ESP) courses, comparing general English with specialized texts.

In this study, based on the notion of ESP (Dudley-Evans & St Johns, 1998; Hutchinson & Waters, 1987), we created a vocabulary list appropriate for use as a learning goal for undergraduates in the agricultural sciences at the University of Miyazaki (UOM). The research questions comprise the following:

1. What are the keywords for semi-popularization articles in agriculture in general?
2. What is the lexical relationship among keywords in the agricultural subdisciplines?

## 2. Materials and methods

### 2.1 Compiling a corpus from semi-popularization articles on agriculture

A one-million-word corpus was compiled from university newsletters and online articles

published by *Science Daily* (<https://www.sciencedaily.com/>) and *Science News for Students* (<https://www.sciencenewsforstudents.org/>). The university newsletters included those published by the 11 U.S. universities whose newsletters on corn production were compiled and analyzed by Muñoz (2015) and those published online by five institutions that have contracted agreements with the UOM. The newsletters and online articles were all published between March 2014 and August 2018.

The articles were collected by the first author and her research collaborators via the Internet and classified into six categories corresponding to the university's six agricultural departments. The collection and classification policies were revised and updated by the raters when necessary. During the process, the target for inter-rater reliability was 1.0, which was finally reached. This process resulted in the compilation of a corpus of 1,179,064 words from semi-popularization articles in the field of agriculture, encompassing its six sub-fields (Table 1). This agricultural corpus (hereafter AG-UOM) was created for teaching English at UOM.

**Table 1: Number of articles and tokens for each subdiscipline**

Department/Subdiscipline	Number of articles	Tokens
Agricultural and Environmental Sciences (AE)	408	274,197
Forest and Environmental Sciences (F)	272	210,286
Biochemistry and Applied Biosciences (B)	191	126,587
Marine Biology and Environmental Sciences (M)	192	118,621
Animal and Grassland Sciences (AG)	238	175,101
Veterinary Sciences (V)	432	273,881
Total		1,179,064

## 2.1 Keywords in AG-UOM

Based on Gilmore and Millar (2018), the AG-UOM was compared to a larger corpus to extract keywords occurring more frequently in the target corpus. Using the Corpus of Contemporary American English (COCA) as a reference, keywords were extracted from the AG-UOM corpus. The relevant COCA subsections (written novels, magazines, and newspaper text) were of the same period as the AG-UOM (2014–2017). The total size of the COCA reference corpus was approximately 48 million tokens. Keyness values for all words in the AG-UOM were calculated using WordSmith Tools, Version 7.0 (Scott, 2016), to indicate the characteristics of each word by log-likelihood.

## 2.2 Coverage in the New General Service List (NGSL), the New Academic Word List

(NAWL), and off-list words

Keywords extracted from the AG-UOM corpus were divided into three groups: 1) Words included in the NGSL (Browne et al., 2013), 2) Words included in the NAWL (Browne et al., 2013), 3) Words excluded in both (1) and (2). The coverage of groups (1)–(3) was calculated in each subdiscipline and compared with the data calculated by Gilmore and Miller (2018). The extracted “off-list” keywords were also compared between subdisciplines, and their uniqueness and commonality were observed.

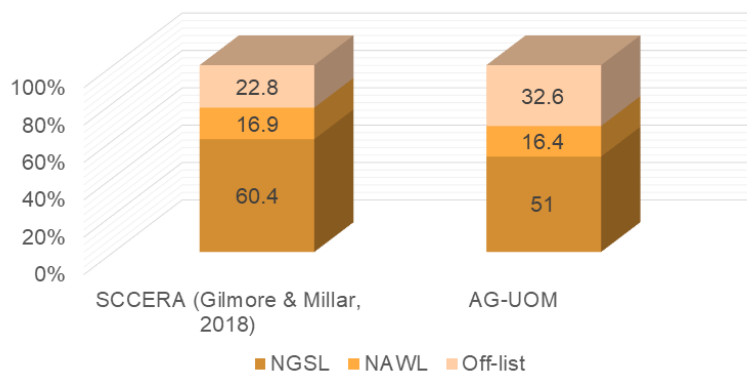
### 2.3 Lexical interrelatedness between departments

Using the data comprising keywords from each subdiscipline, the commonality was measured by means of a hierarchical agglomerative cluster analysis, following the Ward method.

## 3. Results and discussion

### 3.1 AG-UOM corpus keywords

Through comparison with COCA, 1,639 keywords were extracted: 836 were found to occur in the NGSL and 269 in the NAWL, while 534 were found in neither of these two corpora. These off-list keywords can be regarded as lying between general and academic usage in agriculture. Figure 1 shows the coverages of the AG-UOM corpus, and SCCERA (a specialized Corpus of Civil Engineering Research Articles) compiled by Gilmore and Millar (2018). The percentage of off-list keywords was higher in the AG-UOM corpus than the academic corpus SCCERA, while the percentage was almost the same as that of the NAWL. However, the percentage of NGSL in SCCERA was high, suggesting that the AG-UOM corpus contains many characteristic vocabularies.



**Figure 1: Coverage of NGSL, NAWL, and off-list keywords**

### 3.2 Off-list keywords in each agricultural subdiscipline

Figure 1 shows the top 20 off-list keywords in each subdiscipline that represent the characteristics of the agricultural field.

#### Agricultural Environmental Sciences (AE)

**CORN** BEES SOYBEAN WEEDS SOYBEANS GLYPHOSATE KOCHIA  
WEED POLLEN ATRAZINE DICAMBA BEE ANTS KANSAS  
FERTILIZER GROWERS SORGHUM ALFALFA **MOISTURE** ACRE

#### Animal Grasslands Sciences (AG)

BEEF **CALVES** FORAGE **CALF** GRAZING **CALVING** HERD HAY  
LIVESTOCK PASTURE WEANING BULLS SILAGE BULL HEIFERS  
WEANED ACRE KRIS PASTURES **CORN**

#### Biochemistry and Applied Biology (B)

PIRBRIGHT VACCINE INSTITUTE GERMS **MICROBES** **DNA**  
MEALWORMS FMD MICROCYSTIN SPIDERS VACCINES BIOFILMS  
ATMOSPHERIC BIOMASS DR **AVIAN** FMDV INFLUENZA PCR FLU

#### Forest and Environmental Sciences (F)

DEFORESTATION DROUGHT INVASIVE EROSION PERMAFROST  
FORESTRY SERICEA **ECOSYSTEM** ASH TIMBER WILDLIFE CFIA  
CANOPY **ECOSYSTEMS** **MOISTURE** PEAT BIOMASS ERUPTION  
**MICROBES** RAINFOREST

#### Marine Biology and Environmental Sciences (M)

CORAL REEFS MARINE CORALS REEF ALGAE WHALES SHARKS  
**ECOSYSTEMS** BLEACHING WHALE PHOSPHORUS **MAMMALS**  
AQUACULTURE EEL DOLPHINES UNDERWATER FISHERIES  
ZOOPLANKTON **ECOSYSTEM**

#### Veterinary Sciences (V)

**MAMMALS** VETERINARY **DNA** MOSQUITOES CHIMPANZEES  
PRIMATES EXTINCT **CALVING** LEMURS EXTINCTION BUGS  
PRIMATE ELEPHANTS **CALVES** GORILLAS DINOSAURS POULTRY  
**AVIAN** GORILLA FEATHERS

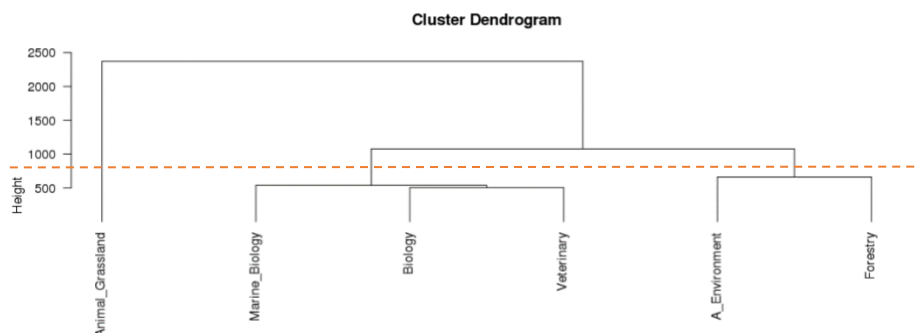
**Figure 2: Top 20 off-list keywords in agriculture subdisciplines**

Not only does each department have its own characteristic vocabulary but there are words that also occur across two departments (CORN in AE and AG; CALF and CALVES in AG and V; DNA in B and V), suggesting that several departments share some research objects and methodologies.

### 3.3 Lexical commonality

The extracted off-list keywords related to the six agricultural

departments/subdisciplines were analyzed with hierarchical agglomerative clustering (Figure 3).



**Figure 3: Cluster analysis results on six subdisciplines**

Two categories emerged: “science for plants” and “science for animals.” However, AG stands on its own, with no great similarity to either of the two clusters.

#### 4. Conclusion

The AG-UOM corpus compiled from semi-popularization articles includes unique vocabulary items, which are commonly used in English speaking countries, especially in agriculture-related contexts. However, these items are not found in the NGSL and NAWL. This study thus identified the appropriate vocabulary for undergraduate students of agriculture who learn English as a foreign language. Further, the cluster analysis suggested lexical relatedness between subdisciplines/departments. Producing word-lists for each discipline with distinctiveness and commonality could provide optimal materials, enabling students to be successful in their specialisms.

#### Acknowledgment

This work was supported by KAKEN 16K02844.

#### References

Browne, C., Culligan, B., & Phillips, J. (2013) *The new general service list*. Available at: <http://www.newgeneralservicelist.org>. (Accessed 29 May 2019).

Chen, Q. and Ge, G. (2007) ‘A corpus-based lexical study on frequency and distribution of Coxhead’s AWL word families in medical research articles (RAs)’, *English for Specific Purposes*, 26(4), pp. 502–514. doi: 10.1016/j.esp.2007.04.003.

Dudley-Evans, T. and St Johns, M.-J. (1998) *Developments in English for specific purposes*. Cambridge University Press.

Gilmore, A. and Millar, N. (2018) 'The language of civil engineering research articles: A corpus-based approach', *English for Specific Purposes*. Elsevier Ltd, 51, pp. 1–17. doi: 10.1016/j.esp.2018.02.002.

Hutchinson, T. and Waters, A. (1987) *English for specific purposes: a learning-centred approach*. Cambridge University Press.

Hyland, K. and Tse, P. (2007) 'Is there an "academic vocabulary"?', *TESOL Quarterly*, 41(2), pp. 235–253. doi: 10.1002/j.1545-7249.2007.tb00058.x.

Martínez, I. A., Beck, S. C. and Panza, C. B. (2009) 'Academic vocabulary in agriculture research articles: A corpus-based study', *English for Specific Purposes*, 28(3), pp. 183–198. doi: 10.1016/j.esp.2009.04.003.

Muñoz, V. L. (2015) 'The vocabulary of agriculture semi-popularization articles in English: A corpus-based study', *English for Specific Purposes*. Elsevier Ltd, 39, pp. 26–44. doi: 10.1016/j.esp.2015.04.001.

Scott, M. (2016) *WordSmith tools version 7.0, Stroud: Lexical analysis software*. Available at: <http://www.lexically.net/wordsmith/version6/>. (Accessed 5 June 2019).

Wang, J., Liang, S. and Ge, G. (2008) 'Establishment of a Medical Academic Word List', *English for Specific Purposes*, 27(4), pp. 442–458. doi: 10.1016/j.esp.2008.05.003.

Yang, M. N. (2015) 'A nursing academic word list', *English for Specific Purposes*. Elsevier Ltd, 37(1), pp. 27–38. doi: 10.1016/j.esp.2014.05.003.