

CORPORA IN EAP

Iztok Kosem

Trojina, Institute for Applied Slovene Studies
Ljubljana, Slovenia

Contact: iztok.kosem@trojina.si

English for Academic Purposes (EAP)

- Origins of EAP around 1972 (SELMOUS Special English Language Materials for Overseas University Students)
 - ▣ Birmingham, Manchester, Newcastle, Leeds
- The term EAP used first in October 1974 (Tim Johns)
- EAP emerged for ESP
 - ▣ EAP “refers to language research and instruction that focuses on the specific communicative needs and practices of particular groups in academic contexts” (Hyland & Hamp-Lyons, 2002)
 - ▣ ESP focuses on tailoring instruction to specific purposes, i.e. the needs of a particular discipline

EAP (continued)

- EAP - Initially, main focus on non-native speakers
- EAP associated mostly with pre-sessional and in-session courses
- The need for EAP increased with increasing numbers of students
 - ▣ 600 % increase in student numbers in the four largest English-speaking countries between 1955 and 2007

EAP and corpora

- Academic texts in general corpora
 - ▣ BNC: 15,8% (16M words)
 - ▣ Bank of English, 2000: 6M words of US academic textbooks
 - ▣ Oxford English Corpus (OEC): included, no figures offered
 - ▣ Problems:
 - Incomplete texts
 - Poor representativeness

EAP and corpora

- Types of EAP corpora used:
 - ▣ Corpora of expert writing (and speech)
 - ▣ Learner corpora
 - Corpora of university student writing
 - ▣ Spoken corpora

Corpora of expert writing and speech

- Jiaotong Daxue English for Science and Technology (JDEST) corpus
 - ▣ The first EAP corpus (1985)
 - ▣ 4M words (originally 1M)
 - ▣ Written only – textbooks and academic works from 10 disciplines (mainly applied sciences disciplines)

- MicroConcord (Academic)
 - ▣ 1M words
 - ▣ Books and articles
 - ▣ Free online access:
<http://langbank.engl.polyu.edu.hk/corpus/microconcord.html>

Corpora of expert writing and speech

- TOEFL 2000 Spoken and Written Academic Language Corpus (T2KSWAL)
 - ▣ 2.7M words (1.7M written, 1M spoken)
 - ▣ written: textbooks, course packs, course management, campus writing
 - ▣ spoken: class sessions, classroom management office hours, study groups, service encounters

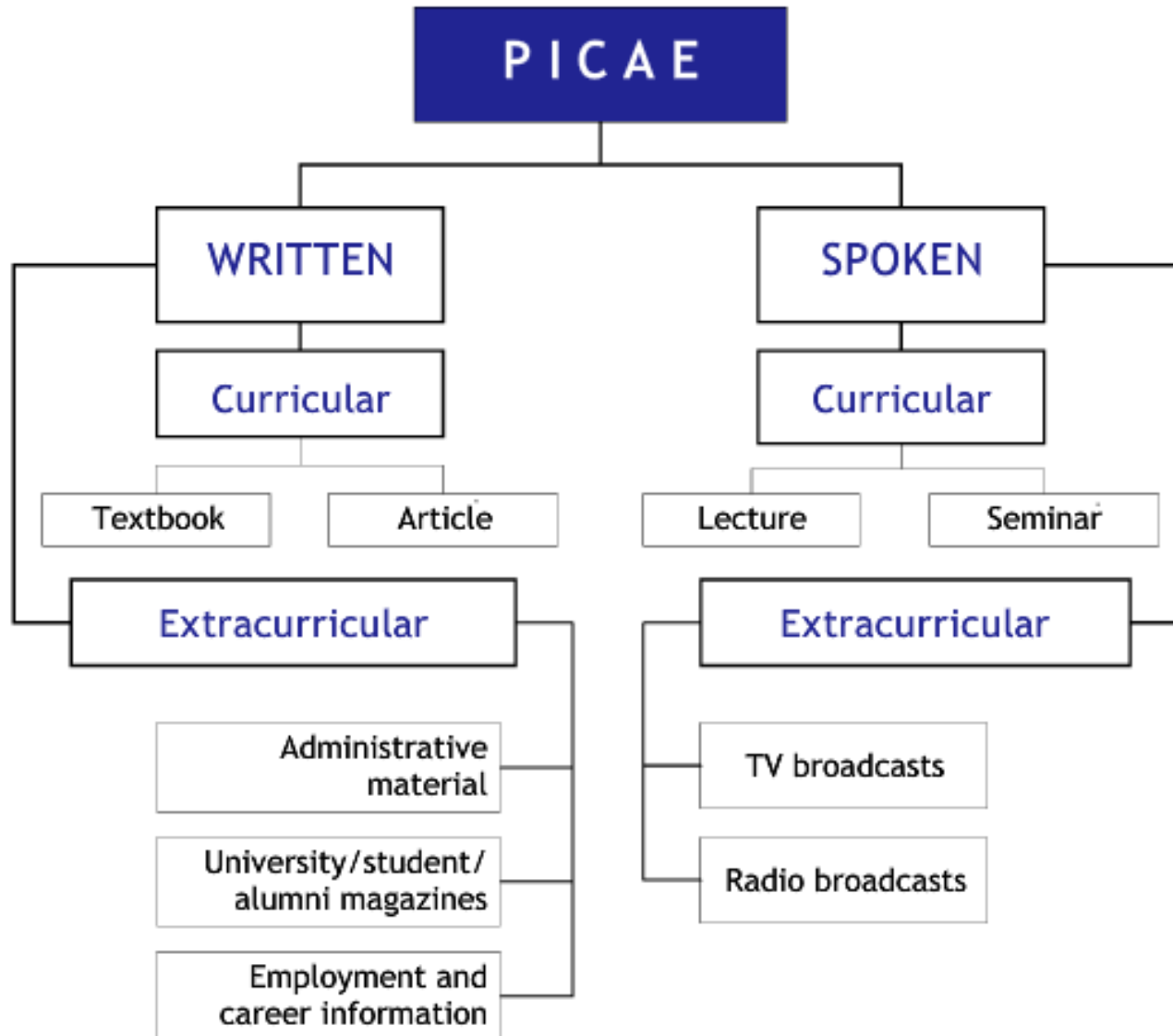
Corpora of expert writing and speech

- The PERC (the Professional English Research Consortium) corpus
 - 17M words
 - journal articles (1995-2002) from 22 science and technology domains
 - 50.000 words per journal
 - Access: <https://scn.jkn21.com/~percinfo/index.html>
(user agreement needs to be submitted)

Corpora of expert writing and speech

- Individual projects (written only)
 - ▣ The Gledhill corpus (Gledhill, 1996; 2000)
 - 0,5M words
 - 150 cancer research articles
 - ▣ The Academic corpus (Coxhead, 2000)
 - 3,5M words
 - 414 articles, textbooks, books from 28 subject areas
 - ▣ The Hyland corpus (Hyland, 2002)
 - 1,4M words
 - 240 journal articles from 8 disciplines
 - ▣ CAJA (Corpus of Academic Journal Articles)
 - 83,5M words
 - 13,116 articles from 28 different disciplines
 - balanced

Corpora of expert writing and speech



Learner corpora

- ICLE (International Corpus of Learner English)
 - ▣ 3,7M words
 - ▣ 16 different mother tongue backgrounds
 - ▣ Essays (“Crime does not pay”, “The role of censorship in Western society”)
- Longman Learners Corpus (LLC)
 - ▣ 10M words
 - ▣ different language backgrounds and levels
- Cambridge Learner Corpus (CLC)
 - ▣ Cca. 35M words
 - ▣ Exam tests
 - ▣ 130 different mother tongues
- **Are these corpora really EAP corpora?**

Learner corpora (academic!)

- Hong Kong University of Science and Technology(HKUST) Learner Corpus
 - ▣ 25M words
 - ▣ Essays from upper-secondary school and university students
 - ▣ available for use in research on a collaborative basis
- BAWE (British Academic Written English)
 - ▣ 6,5M words
 - ▣ Upper-level student writing
 - ▣ Available at Oxford text archive or via Sketch Engine

Learner corpora (academic!)

- MICUSP (Michigan Corpus of Upper-level Student Papers)
 - 2,6M words
 - Upper-level writing (829 papers from 16 disciplines)
 - <http://search-micusp.elicorpora.info/simple/>
- The Uppsala Student English Corpus
 - 1,2M words
 - 1489 essays by 440 Swedish university students of English
 - available at the Oxford Text Archive (<http://www.ota.ahds.ac.uk>)

Spoken corpora

- Michigan Corpus of Academic Spoken English (MICASE)
 - 1,7M words
 - Lectures, meetings, interviews
 - <http://quod.lib.umich.edu/m/micase/>
- British Academic Spoken English (BASE) corpus
 - 1,2M words
 - 160 lectures, 40 seminars
 - Oxford Text Archive, Sketch Engine

Spoken corpora

- The Corpus of English as Lingua Franca in Academic Settings (ELFA)
 - 1 M words
 - 650 non-native speakers (51 L1 backgrounds)
 - 100 EUR for 6 months (50 EUR for each renewal)
- Hong Kong Corpus of Spoken English (HKCSE)
 - 0,9M words
 - 50 hours of lectures, seminars, student presentations,...
 - <http://rcpce.engl.polyu.edu.hk/HKCSE/>

Research

- Research into genres of academic English
 - ▣ Written discourse
 - ▣ Spoken discourse
 - ▣ Academic English vs. general English
- Vocabulary studies
- Phraseology

Research into written genres

□ Research articles

- ▣ Evaluation and stance (e.g. Hunston, 1993; Hyland, 1994, 2001)
- ▣ Grammatical constructions (e.g. Swales et al., 1998)
- ▣ Vocabulary (e.g. Nation, 1990; Coxhead, 2000)
- ▣ Phraseology (e.g. Biber et al., 1999; Swales, 2004)

□ Student writing

- ▣ Undergraduate essay (e.g. Myers, 2001)
- ▣ Research proposal (e.g. Cadman, 2002)
- ▣ PhD thesis (e.g. Swales, 2004; Thompson, 2005)

Research into genres

□ Written

□ BAWE corpus

- 13 genre families identified (50 genres)
- Distinguishing characteristic is discipline
- Some genres can also be subgenres (e.g. literature review)

□ MICUSP

- 7 paper types (argumentative essay, proposal, etc.)
- 8 textual features (abstract, methodology, etc.)

□ Spoken

□ BASE, MICASE

- Several genres identified
- Classroom events (e.g. lectures) and non-classroom events (e.g. office hours)



Research into genres

- Comparing genres
 - ▣ Published genres (e.g. Groom, 2005)
 - ▣ Student writing and academic writing
 - ▣ Student writing and IELTS test writing
- Comparing linguistics features across disciplines in a single genre
 - ▣ *It*-clause hedges in articles from four disciplines (Hewings & Hewings, 2001)
 - ▣ *Stance nouns* in PhD theses (Charles, 2003)
 - ▣ *Sequences of four words* in articles, PhD theses (Hyland, 2008)



Research into genres

□ Comparing genres

▣ Multidimensional analysis (Biber)

- “all spoken registers are similar in their typical linguistic characteristics” (Biber, 2006:223)
- “spoken registers are systematically different from written registers, with respect to a wide range of vocabulary characteristics and lexico-grammatical features”
- Differences between genres of student writing
- Variation among academic registers (disciplines)

Academic English vs. general English

- Longman Grammar of Spoken and Written English (Biber et al., 1999)
 - ▣ Conversation, fiction, newspapers, academic writing
 - ▣ nouns, adjectives, and prepositions are very frequent in academic writing, more than in the other three registers
 - ▣ grammatical features especially characteristic of academic writing are for example nominalizations, stance noun + *of*-phrase, noun phrases with (multiple) modifiers, extraposed *that*-clauses, and relative clauses with the relative pronoun *which*

Academic English vs. general English

- Longman Grammar of Spoken and Written English (Biber et al., 1999)
 - ▣ although verbs and adverbs are not very frequent in academic writing, certain categories of verbs and adverbs are most frequently found in academic writing, such as copula *be*, passive voice, present tense simple, linking adverbials (e.g. *however*)
 - ▣ several grammatical features are frequent only in academic writing (e.g. verbs with inanimate subjects, *that/those* + *of*-phrase, quantifier *each*, coordination tag)

Vocabulary

- students need to understand 95% of the words in the text to gain reasonable comprehension (Nation & Waring, 1997; Nation, 2001)
- English lexicon:
 - High frequency words
 - ACADEMIC VOCABULARY
 - Technical vocabulary
 - Low frequency words

Vocabulary

- List of academic words
 - ▣ University Word List (Xue and Nation, 1984)
 - 836 word families
 - ▣ Academic Wordlist (Coxhead, 2000)
 - Based on 3,5-million-word Academic Corpus
 - 570 word families
 - A word had to occur at least 10 times in each of the top four domain categories

Vocabulary

- List of academic words
 - ▣ Discipline-specific wordlists
 - Medicine (Wang et al., 2008)
 - Engineering (Ward, 2009)
 - Agricultural sciences (Martinez et al., 2009)
 - ▣ Productive wordlist (Paquot, 2007)
 - 838 lemmas

Vocabulary

- Wordlists - weaknesses:
 - corpus data (e.g. old data, unbalanced, small size)
 - excluding 2000 most frequent words
 - word families
 - Word meanings not considered
 - single words, no phraseology
 - Most wordlists not suitable for productive purposes
- Still useful, e.g.:
 - www.uefab.com
 - <http://www.uefab.com/vocab/exercise/awl/ukhe.htm>

Phraseology

- Most studies on lexical bundles and collocations
 - ▣ Lexical bundles (Biber et al., 1999): “a frequently occurring sequence of words which is usually not idiomatic in meaning, and is usually not a complete structural unit”
 - ▣ Important for comprehension/production of academic language
 - ▣ Lexical bundles found in all academic registers, more frequent in spoken ones
 - ▣ Distribution of lexical bundles according to discourse function even more varied (Biber and Barbieri, 2007)

Phraseology

- Lexical bundles
 - Hyland (2008):
 - Lexical bundles in 4 disciplines
 - Over 50% of lexical bundles are discipline-specific
 - Cortes (2004)
 - Students use lexical bundles more frequently and for different purposes than expert writers



Phraseology

□ Collocation


- ▣ Each discipline has many discipline-specific collocation patterns (Gledhill, 1996; Ward, 2007)
- ▣ *It-clauses*: some also discipline specific, different functions in different disciplines (Hewings and Hewings, 2001; Oakey, 2002)
- ▣ Academic Collocation lists (Durrant, 2009; Ackermann et al., 2011)



Phraseology

□ Collocation

- ▣ Analysing collocations in non-native student writing (Howarth, 1996; Granger, 1998)
 - “collocations most common source of learner error”
 - NNS use fewer collocations than NS
 - Underuse of native-like collocations, overuse of atypical collocations
 - BUT atypical collocations found in NS data (Howarth, 1996)

- 
- SKETCH ENGINE (+ the CAJA corpus)

References

- Ackermann, K., Biber, D., Gray, B. (2011). An Academic Collocation List. Corpus Linguistics 2011. Birmingham.
- Biber, D. & Barbieri, F. (2007). Lexical bundles in university spoken and written registers. *English for Specific Purposes* 26(3): 263-286.
- Biber, D., Johansson, S., Leech, G., Conrad, S. & Finegan, E. (1999). *The Longman Grammar of Spoken and Written English*. London: Longman.
- Cadman, K. (2002). English for Academic Possibilities: the research proposal as a contested site in postgraduate genre pedagogy. *Journal of English for Academic Purposes* 1(2): 85- 104.
- Cortes, V. (2004). Lexical bundles in published and student disciplinary writing: Examples from history and biology. *English for Specific Purposes* 23(4): 397-423.
- Coxhead, A. (2000). A New Academic Word List. *TESOL Quarterly* 34(2): 213-238.
- Durrant, P. (2009). Investigating the viability of a collocation list for students of English for academic purposes. *English for Specific Purposes* 28(3): 157-169.
- Gledhill, C. (1996). Science as a Collocation: Phraseology in Cancer Research Articles. In Botley, S., Glass, J., McEnery, T. & Wilson, A. (eds.) *Proceedings of Teaching and Language Corpora 1996* (UCREL Technical Papers Volume 9).
- Gledhill, C. (2000). The Discourse Function of collocation in Research Article Introductions. *English for Specific Purposes* 19(2): 115-135.

References

- Granger, S. (1998b). Prefabricated Patterns in Advanced EFL Writing: Collocations and Formulae. In Cowie, A.P. (ed.) *Phraseology: Theory, Analysis, and Applications*. Oxford: Oxford University Press. 145-160.
- Groom, N. (2005). Pattern and meaning across genres and disciplines: An exploratory study. *Journal of English for Academic Purposes* 4(3): 257-277.
- Hewings, A. & Hewings, M. (2001). Anticipatory it' in academic writing: an indicator of disciplinary difference and developing disciplinary knowledge. In Hewings, M. (ed.) *Academic Writing in Context: Implications and Applications: Papers in Honour of Tony Dudley-Evans*. Birmingham: University of Birmingham Press. 199-214.
- Howarth, P. (1996). *Phraseology in English Academic Writing: Some Implications for Language Learning and Dictionary Making*. Tübingen: Niemeyer.
- Hunston, S. (1993). Evaluation and ideology in scientific writing. In Ghadessy, M. (ed.) *Register analysis: Theory and practice*. London: Pinter. 57-73.
- Hyland, K. (1994). Hedging in academic writing and EAP textbooks. *English for Specific Purposes* 13(3): 239-256.
- Hyland, K. (2001). Humble servants of the discipline? Self-mention in research articles. *English for Specific Purposes* 20(3): 207-226.
- Hyland, K. (2002). Activity and Evaluation: Reporting practices in academic writing. In Flowerdew, J. (ed.) *Academic Discourse*. Harlow: Pearson. 115-130.

References

- Hyland, K. (2008). As can be seen: Lexical bundles and disciplinary variation. *English for Specific Purposes* 27(1): 4-21.
- Hyland, K., and Hamp-Lyons, L. (2002). "EAP: Issues and directions." *Journal of English for Academic Purposes* 1 (1): 1-12.
- Martínez, I.A., Beck, S.C. & Panza, C.B. (2009). Academic vocabulary in agriculture research articles: A corpus-based study. *English for Specific Purposes* 28(3): 183-198.
- Myers, G. (2001). 'In my opinion': the place of personal views in undergraduate essays. In Hewings, M. (ed.) *Academic Writing in Context : Implications and Applications : 365 Papers in Honour of Tony Dudley-Evans*. Birmingham: University of Birmingham Press. 63-78.
- Nation, P. (1990). *Teaching and Learning Vocabulary*. Boston: Heinle & Heinle.
- Nation, P. (2001). *Learning Vocabulary in Another Language*. Cambridge: Cambridge University Press.
- Nation, P. & Waring, R. (1997). Vocabulary size, text coverage and word lists. In Schmitt, N. & McCarthy, M. (eds.) *Vocabulary: Description, Acquisition and Pedagogy*. Cambridge: Cambridge University Press. 6-19.
- Oakey, D.J. (2002). Formulaic language in English academic writing: A corpus-based study of the formal and functional variation of a lexical phrase in different academic disciplines. In Reppen, R., Fitzmaurice, S. & Biber, D. (eds.) *Using Corpora to Explore Linguistic Variation*. Philadelphia: John Benjamins. 111-129.

References

- Paquot, M. (2007). Towards a Productively-oriented Academic Wordlist. In Walinski, J., Kredens, K. & Gozdz-Roszkowski, S. (eds.) *Corpora and ICT in Language Studies: PALC 2005*. Frankfurt am Main: Peter Lang. 127-140.
- Swales, J. (2004). *Research Genres: Exploration and Applications*. Cambridge: Cambridge University Press.
- Swales, J., Ahmad, U.K., Chang, Y., Chavez, D., Dressen, D.F. & Seymour, R. (1998). Consider this: The role of imperatives in scholarly writing. *Applied Linguistics* 19: 97- 121.
- T2K-SWAL: Biber, D., Conrad, S., Reppen, R., Byrd, P. & Helt, M. (2002). Speaking and Writing in the University: A Multidimensional Comparison. *TESOL Quarterly* 36(1): 9-48.
- Thompson, P. (2005). Points of focus and position: Intertextual reference in PhD theses. *Journal of English for Academic Purposes* 4(4): 307-323.
- Wang, J., Liang, S.-l. & Ge, G.-c. (2008). Establishment of a Medical Academic Word List. *English for Specific Purposes* 27(4): 442-458.
- Ward, J. (2009). A basic engineering English word list for less proficient foundation engineering undergraduates. *English for Specific Purposes* 28(3): 170-182.
- Xue, G.Y. & Nation, P. (1984). A University Word List. *Language Learning and Communication* 3: 215-229.