

Advances in Corpus-Informed ESP Research and Teaching

Laurence Anthony

Center for English Language Education in Science and Engineering (CELESE),
Waseda University, Japan
anthony@waseda.jp
http://www.antlab.sci.waseda.ac.jp



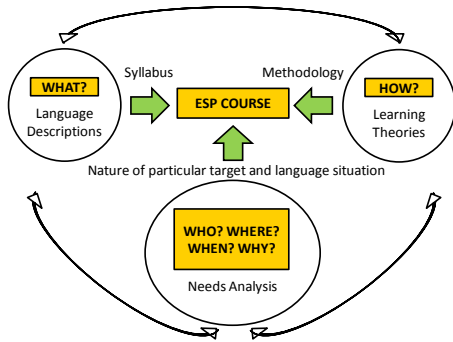
Faculty of Science and Engineering, Waseda University

Outline

- ESP Course Design
- ESP needs analysis
 - Necessities, Lacks, and Wants
 - Problems in ESP
- Traditional approaches to ESP Program Design
- Applying corpora and corpus tools in ESP research
 - Deciding what to teach
- Applying corpora and corpus tools in ESP teaching
 - Deciding how to teach it
- A New Proposal for ESP Program Design

2

ESP Course Design: Hutchinson & Waters (1987)



3

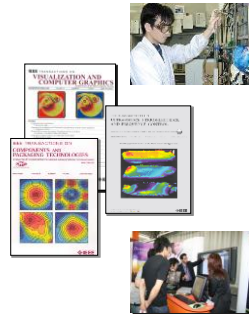
ESP Needs Analysis: Who? Where? When? Why?

- **Necessities:** What do the learners *need* to learn to achieve the goal(s) of the course?
- **Lacks:** Which of the necessities do the learners *lack* at present?
- **Wants:** What do the learners *want* to learn?
- **Factors**
 - text types (e.g., research papers, essays, specifications, instructions)
 - skills (e.g., reading, listening, speaking, writing, fluency)
 - topics/themes (e.g., science, engineering, business, medicine)
 - language (e.g., grammar, vocabulary, phrases, pronunciation)

4

ESP Needs Analysis: Problems in ESP

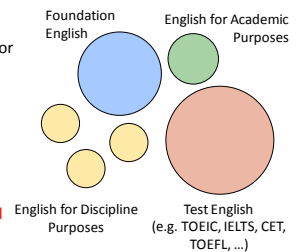
- Most students' ESP *future* needs are highly *specific* (see Hyland, 2002, 2004)
 - e.g. **technical writing** and **presentation skills** in STEM disciplines
- Developing these skills is resource intensive
 - small class sizes
 - experienced instructors
 - funding
 - time



5

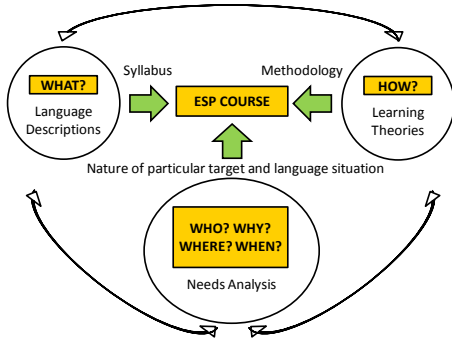
A solution to ESP problems?

- ESP courses in a traditional English program
 - give ESP courses **elective** or **non-credit** status
 - introduce **strict entry requirements**
 - teach ESP courses only to **interested specialist departments**
 - offer only **short-term** ESP courses based on **external funding**
 - position ESP courses on the **fringes** of the English program (Anthony, 2009)



6

ESP Course Design: Hutchinson & Waters (1987)



7

What language should we teach?



"We need to develop a new ESP course for physicists."

"ESP must involve teaching the literacy skills which are appropriate to the purposes and understandings of particular academic and professional communities."

(Hyland, 2002: 385)



8

What language should we teach?

A collage of images related to physics. It includes a portrait of Albert Einstein, a diagram of quantum theory, a book cover for 'The Physics of Quantum Theory', a graph of world population changes, and a student looking at a whiteboard with physics equations.

9

What language should we teach?



vocabulary?

phraseology?

tense/voice?

citation patterns?

discourse structure?

discipline variation?

???

10

Top 10 most cited articles in the *ESP Journal*

- Hyland, K. (2008). As can be seen: **Lexical bundles** and **disciplinary variation**.
- Rogerson-Revell, P. (2007) Using English for **International Business**: A European case study.
- Mudraya, O. (2006). **Engineering English**: A **lexical frequency** instructional model
- Pecorari, D. (2006). Visible and occluded **citation features** in **postgraduate second-language writing**
- Matsuda, P.K., Tardy, C.M. (2007). **Voice in academic writing**: The rhetorical construction of author identity in blind manuscript review
- Lee, D., Swales, J. (2006). A corpus-based EAP course for **NNS doctoral students**: Moving from available specialized corpora to self-compiled corpora
- Charles, M. (2006). **Phraseological patterns** in **reporting clauses** used in **citation**: A corpus-based study of theses in **two disciplines**
- Cheng, A. (2006). Understanding learners and learning in ESP genre-based **writing instruction**
- Ozturk, I. (2007). The **textual organisation** of **research article** introductions in applied linguistics: **Variability within a single discipline**
- Bhatia, V.K. (2008). **Genre analysis, ESP and professional practice**



11

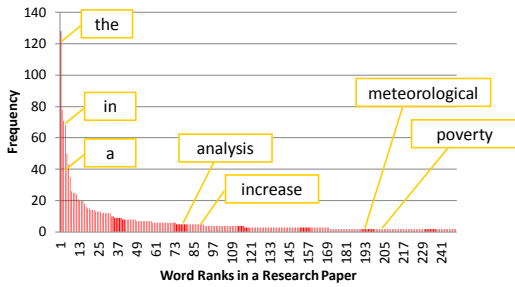
Corpus-Informed ESP research: Definition of Corpus Linguistics (Biber, 1998)

- It is an **empirical** (experimental) approach
 - An analysis of actual patterns of use in target texts
- It uses a **corpus** of natural texts as the basis for analysis
 - Corpus = a representative sample of target language stored as an electronic database
- It relies on **computer software** for analysis
 - Results are generated using automatic and interactive techniques
- It depends on both **quantitative** and **qualitative** analytical techniques
 - Observations are counted and results are interpreted



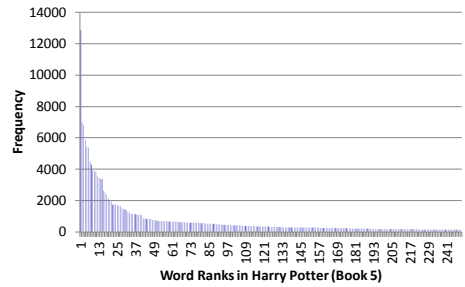
12

Corpus-Informed ESP research: The vocabulary of physics



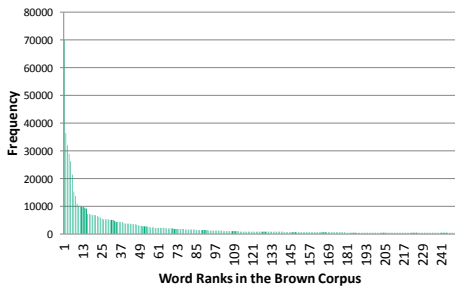
13

Corpus-Informed ESP research: The vocabulary of novels



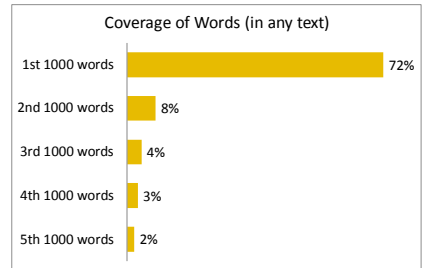
14

Corpus-Informed ESP research: The vocabulary of general English



15

Corpus-Informed ESP research: The vocabulary of (any) English



Based on Brown Corpus

16

Corpus-Informed ESP research: Teaching vocabulary

"It is important that learners have access to lists of high-frequency and academic words and are able to obtain frequency information from dictionaries." (p. 219)

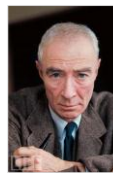
"Priority should be given to high-frequency words and to words that clearly fulfill language use needs." (p. 303)



P. Nation (2001)

17

Corpus-Informed ESP research: Technical terms in physics (and other fields)



satellite

image

particle

diagram

???

18

Corpus-Informed ESP research: Technical terms in physics (and other fields)

Rank	Physics	Math	Biology	Chemistry	Comp. Sci.
1	population	record	cells	reaction	fault
2	satellite	system	skin	solution	cache
3	census	solution	cell	mmol	computer
4	data	equations	expression	mol	algorithm
5	nighttime	model	mice	bond	is
6	countries	solutions	were	structure	number
7	urban	nonlinear	protein	observed	node
8	dmsp	theorem	induced	spectra	systems
9	changes	equation	keratinocytes	energy	performance
10	ngdc	stability	tumor	complexes	computers

Highest Ranked Keywords in Discipline-Specific Corpora

19

Corpus-Informed ESP research: Phraseology in physics (and other fields)



"... can be seen in ..."

"... reported in the literature"

"... changes in light intensity"

"... evaluate the effectiveness of ..."

20

Corpus-Informed ESP research: Phraseology in physics (and other fields)

Rank	Biology	Electrical Engineering	Business studies	Applied Linguistics
1	in the presence of	on the other hand	on the other hand	on the other hand
2	in the present study	as shown in figure	in the case of	at the same time
3	on the other hand	in the case of	at the same time	in terms of the
4	the end of the	is shown in figure	at the end of	on the basis of
5	is one of the	it can be seen	on the basis of	in relation to the

Hyland, K. (2008:12). As can be seen: Lexical bundles and disciplinary variation. *English for Specific Purposes*.

21

Corpus-Informed ESP research: The voice of physics (and other fields)

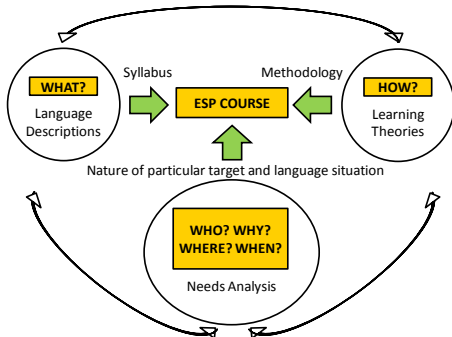


"We show that ..."

⇒ "It was shown that ..."?

22

ESP Course Design: Hutchinson & Waters (1987)



23

How should we teach ESP?

"Numerous studies now show the extent to which language features are specific to particular disciplines, and that the best way to prepare students for their studies is not to search for universally appropriate teaching items, but to provide them with an understanding of the features of the discourses they will encounter in their particular courses.."

(Hyland, 2008: 20)



24

How should we teach ESP?

- The language of specialist subjects is **highly variable** (Hyland, 2002; Hyland, 2004; Hyland and Bondi, 2006; Paltridge, 2009; Biber, 1992; Lea, 1996)
- But, this does **NOT** mean we should identify and teach the unique features of a discipline in the ESP classroom
- We need to teach learners about **probabilistic variation** in core elements (Anthony, 2012)
 - ESP teachers need to help learners understand
 - what features vary, how features vary, when features vary
 - ESP teachers need to help learners
 - recognize, analyze, and estimate **probabilistic variation** in language features across texts and genres

25

Corpus-Informed ESP teaching: Teaching about probabilistic variation

- How can ESP teachers help learners understand **what, how and when** language features vary in and across different disciplines (and genres)?
- How can ESP teachers empower students to be able to identify **what, how and when** language features vary in future (unseen) texts?
 - ⇒ Introduce **Data-Driven Learning (DDL)** into the ESP classroom

26

Corpus-Informed ESP teaching: Teaching about probabilistic variation

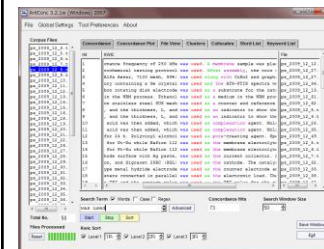
- Characteristics of Data Driven Learning (DDL):
 - A focus on the exploitation of **authentic materials**
 - A focus on **real, exploratory tasks and activities**
 - A focus on **learner-centered activities**
 - A focus on the use and **exploitation of tools**



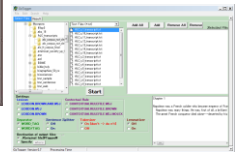
(Bernd Rüschoff, 2010)

27

Corpus-Informed ESP teaching: Teaching about probabilistic variation



AntConc (ver.3.3.x)



Gotagger (ver. 0.7)

28

Example: Teaching *Biographies* writing



Oliver Woodford received the MEng degree in engineering from Cambridge University in 2002, before spending some time in the embedded processor industry. He is currently studying for the DPhil degree in engineering at Oxford University where he specializes in computer vision and, in particular, prior models for near-view synthesis.



Philip Torr received the DPhil degree from the University of Oxford under Professor David Murray, working there as a research fellow for a further three years, and remains a visiting fellow there. He then worked for six years at Microsoft Research, first in Redmond, Washington, then in Cambridge, United Kingdom, founding the vision side of the Machine Learning and Perception group. He is now a professor in computer vision and machine learning at Oxford Brookes University. He has won several awards, including the Mann prize (the highest honor in vision) in 1998, and is a Royal Society Wolfson Research Merit Award holder. He was involved in the algorithm design for Boko, released by 2023, which has won a number of industry awards. He continues to work closely with this Oxford-based company, as well as other companies such as Sony and Sharp. Recent SIGGRAPH work on VideoTrace with the University of Adelaide has been featured extensively on the Internet, including slashdot. He is a senior member of the IEEE.

29

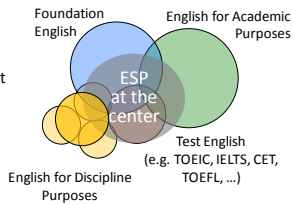
Example: Teaching *Biographies* writing

- **Laurence Anthony** received the M.A. degree in TESL/TEFL, and the Ph.D. in applied linguistics from the University of Birmingham, Birmingham, U.K., and the B.Sc. degree in mathematical physics from the University of Manchester Institute of Science and Technology (UMIST), Manchester, UK. He is a Professor in the Faculty of Science and Engineering at Waseda University, Tokyo, Japan. His primary research interests are in educational technology, corpus linguistics, and natural language processing.

30

A proposal for ESP course/program design

- Put ESP at the **center of program design**
 - integrating all English courses to build ESP skills
 - working closely with subject specialists to provide real-world ESP experiences
 - dividing admin/teaching tasks among full-time and part-time faculty
- Teach students **how to analyze language** in their target disciplines using **corpus tools and methods**



Anthony, L. (2009). "ESP at the center of program design," in K. Fukui, J. Noguchi, & N. Watanabe (Eds.), *Towards ESP bilingualism* (in Japanese) (pp. 18-35). Osaka, Japan: Osaka University Press, 2009.

31

Conclusions

- Corpus approaches have proved to be very effective in ESP research and language teaching
 - They provide researchers with ways to identify common and diverse features of language in and across disciplines
 - They allow teachers and students to identify and measure language variation in and across disciplines
 - They empower students (and teachers) to answer their own questions about specialized English now and in the future
- Corpus approaches cannot be introduced into the classroom in an ad-hoc fashion
 - Technical issues (software/hardware) need to be addressed
 - Carefully designed student materials are needed
 - User (teacher/student) training is essential

32