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

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

ESP Course Design: Hutchinson & Waters (1987)


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

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 (Hyland, 2002)

Foundation English
English for Academic Purposes
English for Discipline Purposes
Text English (e.g. TOEIC, IELTS, CET, TOEFL, ...)
ESP Course Design: Hutchinson & Waters (1987)

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)

Top 10 most cited articles in the ESP Journal


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
Corpus-Informed ESP research: The vocabulary of physics

Corpus-Informed ESP research: The vocabulary of novels

Corpus-Informed ESP research: The vocabulary of general English

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

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

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

Highest Ranked Keywords in Discipline-Specific Corpora

---

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 ...
```

---

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

```
We show that ...
"It was shown that ...?"
```

---

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

```
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)

---

ESP Course Design: Hutchinson & Waters (1987)

```
WHAT?  Language Descriptions
SYLLABUS

WHEN?  Nature of particular target and language situation
NEEDS ANALYSIS

WHERE?  Methodology

HOW?  Learning Theories

ESP COURSE
```

---

How should we teach ESP?
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

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**

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

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.
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

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