Understanding Writing and Oral Presentation
English in Science and Engineering: A Scientific Analysis

Laurence Anthony
Center for English Language Education in Science and Engineering (CELESE), Waseda University (早稲田大学), Japan
anthony@waseda.jp
http://www.antlab.sc.waseda.ac.jp

Outline

- Background
  - English in Asia today
  - Teaching the English needed in Asia today
    - Definition of English for Specific Purposes (ESP)
      - Problems and solutions in ESP for science and engineering
- ESP Course Design
  - Needs, Language descriptions, Learning theories
  - The language of research paper writing
    - general vocabulary, technical terms, phraseology, tense/voice
  - The language of oral presentations
    - vocabulary, style, discourse markers
- Applications of language analysis in the classroom
  - Teaching biography writing and presentation Q&A

Background: English in Asia today

In 2011, 382 employees took language courses totaling 11,352 training hours.

Background: Teaching the English needed in Asia today

- Definition of English for Specific Purposes (ESP)
  (Dudley-Evans, T. & St. John, M. J., 1998: 4-5)
  - ESP is defined to meet specific needs of the learner;
  - ESP makes use of the underlying methodology and activities of the discipline it serves;
  - ESP is centered on the language (grammar, lexis, register), skills, discourse, and genres appropriate to these activities.
Background: Problems in ESP for science and engineering

- Most students’ ESP 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

Background: A solution to the problems in ESP?

- 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)

Background: A real solution to the problems in ESP

- Put ESP at the center of program design
  - integrating all English courses to build ESP skills
  - combining efforts of English teachers and subject specialists to provide effective ESP experiences
  - needs analysis
  - materials design
  - teaching practices
  - student evaluation

ESP Course Design: Hutchinson & Waters (1987)

- WHAT? Language Descriptions
- HOW? Learning Theories
- WHO? WHERE? WHEN? WHY?


- Necessities: What do the students need to learn to achieve the goal(s) of the course?
- Lacks: Which of the necessities do the students lack at present?
- Wants: What do the students want to learn?

Factors

- topics/themes (e.g., science, engineering, business, medicine)
- skills (e.g., reading, listening, writing, speaking, fluency)
- text types (e.g., research papers, essays, specifications, presentations)
- language (e.g., grammar, vocabulary, phrases, pronunciation)
**ESP Language Descriptions: What language should we teach?**

**Corpus-informed language analysis**

- **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 (plural: corpora) = 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 Development Process**

1. **Stage 1: Design a corpus**
   - Choose a target area of language use
   - Review the literature on language features
   - Search for pre-built corpora in target area
   - Empirical investigation
   - Design your own corpus (DIY)
   - Decide a sampling procedure
   - Search for texts and save as TEXT (Annotate the corpus)

2. **Stage 1: Design a corpus**
   - Random sampling
   - Stratified sampling
Corpus-informed language analysis
Stage 1: Design a corpus (whole population)

Corpus-informed language analysis
Stage 2: Choose a software tool

Tools used to analyze corpora

- AntConc
- Longman Mini-concordancer
- MonoConc Pro
- Oxford Concordancing Program
- Sarah (with BNC)
- Sketch Engine
- WMatrix
- Wordsmith Tools
- Xaira (with BNC XML or your own)
- Other


Tools used to analyze corpora

- WordSmith Tools (Scott, M., 2012)
- AntConc (Anthony, L., 2012)
- COCA (Davies, M., 2012)
- BNC Web (Hoffmann et al., 2012)
- AntConc (Anthony, L., 2012)

Corpus-informed language analysis
Stage 2: Choose a software tool

- Freeware
- Multiplatform
  - Win 95/98/NT/XP/7
  - Linux
  - OS X
- Single-file portable app
- Unicode compliant
- HTML/XML tag handing
- Search Features
  - Wildcard/Regex
- Tools
  - KWIC Concordancer
  - Distribution Plot
  - File View
  - Clusters/N-grams
  - Collocates
  - Word Frequency
  - Keyword Frequency

Corpus-informed language analysis
Stage 3: Analyze your data!
Corpus-Informed Language Analysis: The vocabulary of physics research papers

- satellite
- image
- particle
- diagram
- ??

Corpus-Informed Language Analysis: The vocabulary of novels

Corpus-Informed Language Analysis: The vocabulary of general English

Corpus-Informed Language Analysis: 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)
Corpus-Informed Language Analysis: 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   | ngst        | stability | tumor | complexes | computers
```

Highest Ranked Keywords in Discipline-Specific Corpora

Corpus-Informed Language Analysis: Phraseology in physics (and other fields)

```
"... it is important that ..."
"... can be seen in ...
"... in this paper ...
"... we found that ...
```


Corpus-Informed Language Analysis: The voice of physics (and other fields)

```
"We show that ...
"It was shown that ...
```
ESP Language Descriptions (Presentations):
What language should we teach?

- vocabulary?
- phraseology?
- style?
- hedging?
- Q&A?
- ???

Corpus-Informed Language Analysis:
Keywords in technical presentations (MIT OCW)

Corpus-Informed Language Analysis:
The use of "slide", "I", and "you" in presentations

Corpus-Informed Language Analysis:
Introducing the next slide in presentations

- Common phrases using "slide"
  - "... on the next slide I show you ..."
  - "... and the next slide is a bubble chamber ...
  - "... the next slide shows you a corona discharge ...
  - "... the next slide shows you something similar ...
  - "... the next slide shows you the tunnel ...

- Common phrases using "show"
  - "... I showed you a picture ...
  - "... I have just shown you ...
  - "... I'll show you here ...
  - "... I'll show you that ...
  - "... Let me show you a case where ...

Corpus-Informed Language Analysis:
Presentation style

- Research paper style (not good in presentations)
  - Results
  - Western Blot Analysis
    - Using Western blot analysis, GC-17 was demonstrated to specifically recognize two commercially available recombinant ER-s proteins and show no cross-reactivity to an ER-(alpha) recombinant protein (Figure 3; A, B, and C).
  - The recombinant short-form ER-s protein (RP311) from Affinity Bioreagents Inc. has an estimated molecular size of 53 kDa and represents a polypeptide (corresponding to amino acid residues 43 to 530) translated from the second initiation codon of the ER-B transcript.

Corpus-Informed Language Analysis:
Presentation style

- Better ...
  - Results
    - Let me now show you the results
    - Western Blot Analysis
      - First, I'll describe the Western Blot Analysis.
      - Using Western blot analysis, GC-17 was demonstrated to specifically recognize two commercially available recombinant ER-s proteins and show no cross-reactivity to an ER-(alpha) recombinant protein (Figure 3; A, B, and C).
      - What we found was that GC-17 recognized two commercially available recombinant ER-s proteins.
      - It also showed no cross-reactivity to an ER-(alpha) recombinant protein.
      - You can see this here.
Corpus-Informed Language Analysis: Hedging in presentations

- Hedging in presentations ("I" - strong possibility)
  - "... I can almost predict when it happens ..."
  - "... I can always do that ..."
  - "... so now I can predict that ...
  - "... It can actually be much better ...
- Hedging in presentations ("We" - account of previous actions)
  - "... we could change L in the circuit ...
  - "... we could solve this ...
- Hedging in presentations ("You" - abstract example)
  - "... you could also think of it as ...
  - "... you could imagine that running out ...
- Hedging in presentations ("It" - hypothetical case)
  - "... it may be like this ...
  - "... it may not melt at all ...

ESP Learning Theories: How should students learn language?

"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 Learning Theories: How should students learn language?

- The language of specialist subjects is highly variable (Hyland, 2002; Hyland, 2004; Hyland and Bondi, 2006; Patridge, 2009; Biber, 1992; Lea, 1996)
- But, a student’s specialist area is very likely to change through his or her career
- Students need to know about probabilistic variation in core language elements (Anthony, 2012)
  - Students need to understand what features vary, how features vary, when features vary
  - Students need to recognize, analyze, and estimate probabilistic variation in language features across texts and genres

The ESP Specificity Continuum


- General ESP (e.g., academic listening, note-taking, logical structures, visualizing data)
- Narrow ESP (e.g., research article writing, presentations)
- Specific ESP (e.g., nuclear physics terminology, reactor safety manuals)

Corpus-Informed ESP Classroom Practices: Learning about probabilistic variation

- How can ESP teachers help students 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 Classroom Practices: Learning 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

Example: Teaching Biographies writing

Example: Teaching Biographies writing

Example: Teaching Biographies writing
Example: Teaching Biographies writing

"A short biography and passport-style photograph of every author should be provided for Survey Papers, Papers and Brief Papers when requested by the Editor."

Elsevier Automatica
http://www.elsevier.com/wps/find/journaldescription.cws_home/270/authorinstructions

"Include in the manuscript a short (maximum 50 words) biography of each author"

Elsevier Pattern Recognition
http://www.elsevier.com/wps/find/journaldescription.cws_home/328/authorinstructions

"Biographies should be brief."

IEEE Canada
http://www.ieee.ca/journal/authorinfo_style.html

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.

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.

Example: Dealing with Presentation Q&A Sessions

- Why is Q&A so difficult?

  "The questioner had really poor English and so couldn't understand what he was saying."  
  "The questioner was really confusing. She kept changing the words of her question, and repeating herself."  
  "The questioner didn't ask a question at all. He just made a comment and that was it."

  "The audience was really hostile. They were just trying to find problems in my work."
Example: Dealing with Presentation Q&A Sessions

- Typical questions asked by a presentation audience
  - Background
    - What are you doing? What is your aim?
    - What problem are you trying to solve?
  - Body
    - What is the meaning/value of your data?
    - How/Why did you decide your method?
    - What does a word/number/table/chart mean?
  - Discussion
    - How can the results be applied?
    - How will you continue your research from now?

Example: Dealing with Presentation Q&A Sessions

- Understanding the question
  - Questions are not usually direct.
    - "Why did you perform that experiment?"
  - Questions follow the pattern ...
    - Frame → Issue → Request
      - "At the beginning of the presentation, ..."
      - "... you showed a photograph of the proposed robot."
    - Frame → Issue → Comment
      - "Did you explain how big it was?"

Example: Dealing with Presentation Q&A Sessions

- Typical questions asked by an Asian audience
  - WH: How big was it? (50%)
  - Comment: I think this is quite big. (20%)
  - Indirect: Could you tell me how big it was? (10%)
  - YN: Was it big? (10%)
  - A or B: Was it big or small? (10%)

Example: Dealing with Presentation Q&A Sessions

- Typical questions asked by a Native English audience
  - Indirect: Could you tell me how big it was? (31%)
  - YN: Was it big? (24%)
  - WH: How big was it? (21%)
  - Comment: I think this is quite big. (14%)
  - A or B: Was it big or small? (3%)
  - Negative: Wasn’t it big? (3%)
  - Rhetorical: How big is it? (3%)

Example: Dealing with Presentation Q&A Sessions

- Conclusions
  - English for Specific Purposes (ESP) is a useful approach for learning the language needed by scientists and engineers
    - It is based on real student needs
    - It is based on analyses of actual language usage
  - Corpus tools are being increasingly used by both teachers and students of science and engineering
    - They provide teachers and students with ways to identify the language of specialist disciplines
    - They empower teachers and students to answer questions about specialized English without needing native-speaker support