Data-Driven Learning (DDL) in the Technical Writing Classroom

Laurence Anthony
Center for English Language Education in Science and Engineering (CELESE), Waseda University, Japan
anthony@waseda.jp
www.laurenceanthony.net
@antlabjp
Plan of workshop

- **Part 1: Basic principles**
  - Technical writing course design
  - Data-Driven Learning (DDL)
  - Corpus data sources and software tools

- **Part 2: DDL in the technical writing classroom**
  - building of custom, discipline-specific corpora
  - pre-course target language identification
  - in-course uses of corpora and corpus tools
  - introducing DDL in a full-scale writing program

- **Part 3: Discussion**
  - Challenges and limitations of DDL
  - Future work in DDL
Part 1: Basic principles
course design; DDL; corpus data sources and software tools
(ESP) Course design
Adapted from Hutchinson & Waters (1987)

WHAT? Language Descriptions

Syllabus

(ESP) COURSE

Methodology

HOW? Learning Theories

WHO? WHERE? WHEN? WHY?

Needs Analysis

Nature of particular target and language situation
(ESP) Course design
Anthony (2018)

Introducing English for Specific Purposes
Laurence Anthony
Routledge Introductions to English for Specific Purposes
http://www.routledge.com/9781138936652
Technical writing course design

Challenges

- **Varying needs**
  - cross-departmental (interdisciplinary) courses
  - varying levels of discipline-specific knowledge and experience
  - varying levels of English ability

- **Varying language and learning objectives**
  - language, skills, and genres of specialized disciplines vary significantly (Biber, 1988; Hyland, 2002, 2008; Hyland & Bondi, 2006)

- **Varying materials and methods**
  - teacher-centered vs learner-centered?
  - presentation → practice → production?
    - prescriptive vs inductive? task-based? Data-driven-learning (DDL)
Technical writing course design
Challenges (specialist knowledge dilemma)

Which of the following sentences have mathematicians used in journal articles?

- "Let us suppose that this occurs."
- "It is easy to see that the equation is a smooth function."
- "We have a funny feeling about this result."
- "Of course, we have no reason to believe that this picture is complete."
- "Why was the Fibonacci sequence used?"
- "But if the initial tumor is large enough, then it will grow."
- "The graph is clearly weird."
- "Here, you note that the condition above is satisfied."
- "Anyway, we have that EQT is bounded"
Data-Driven Learning (DDL)
A definition

- **Boulton (2012: 263)**
  - DDL is any use of a language corpus by second or foreign language users

- **Ruschoff, B. (2010)**
  - Characteristics of DDL...
    - a focus on real, exploratory tasks and activities
    - a focus on the exploitation of authentic materials
    - a focus on the use and exploitation of tools
    - a focus on learner-centered activities

Data-Driven Learning (DDL)

A definition

- **Learners as language 'researchers'** (Johns, 1991)
  - analyze the corpus with software tools
  - observe bottom-up, lexico-grammatical features and top-down rhetorical and discourse features
  - identify patterns, deduce rules, form hypotheses
  - apply findings in future receptive or productive tasks

- **Teachers as 'facilitators'** show learners how to...
  - collect corpus data
  - select suitable corpus software tools
  - form search queries
  - interpret results
  - apply what is discovered
Data-Driven Learning (DDL)
Advantages with using a data-driven approach

- Teachers do not need to be experts in non-English subjects
  - Teachers need only teach the DDL approach to students

- Students can learn both general and unique features of English in their specialized field
  - Most English textbooks only cover general English

- Students learn life-long skills
  - Students need only learn how to create and analyze corpora of a target text type (e.g. research papers, presentations, emails, ...)

Data-Driven Learning (DDL)

Selected research on DDL

- DDL is successful with older (or adult) learners that have developed a basic knowledge of general English
  - Johansson (2009); Kennedy & Miceli (2001)

- DDL is useful for learning grammar basics in the beginner level EFL classroom
  - Chujo & Oghigian (2008; 2012)

- DDL approaches result in large overall effects for both control/experimental group comparisons (d = 0.95) and for pre/posttest designs (d = 1.50)
  - Boulton & Cobb (2017)

Corpus data sources and software tools
The most popular corpus tools

"Which computer programs do you use for analysing corpora?"
International survey of corpus linguists. Responses: 891. (Tribble, 2012)
Corpus data sources and software tools

Online corpora (with integrated tools)

- corpus.byu (COCA, COHA, BNC, GloWbE, Time Magazine, Soap Opera, ...)
  - corpus.byu.edu/
- CQPweb (multiple corpora)
  - cqpweb.lancs.ac.uk/
- BAWE, BASE, Brown, ...
  - ca.sketchengine.co.uk/open/
- BNCweb
  - bncweb.lancs.ac.uk/bncwebSignup/user/login.php
- MICUSP
  - micusp.elicorpora.info/
Corpus data sources and software tools

Offline corpora (requiring dedicated corpus tools)

- Australian Corpus of English (ACE)
- British Academic Written English (BAWE) Corpus
- British National Corpus (BNC)
- Brown Corpus (+FROWN)
- Business Letter Corpus
- Corpus of Contemporary American English (COCA)
- Enron Email Dataset
- Lancaster Olso Bergen (LOB) Corpus (+ FLOB, +BE06)
- Michigan Corpus of Upper-Level Student Papers (MICUSP)
- Open American National Corpus (OANC)
- PolyU Business Corpora
- SRI American Express travel agent dialogue corpus
- The Twitter Political Corpus
- Wellington Corpus of Written NZ English (WWC)
- ...
Corpus data sources and software tools
Lists of corpora

- martinweisser.org/corpora_site/CBLLinks.html
- courses.washington.edu/englhtml/engl560/corplingresources.htm
- linguistics.stanford.edu/resources/corpora/corpus-inventory
- www.helsinki.fi/varieng/CoRD/corpora/index.html
- www.unc.edu/~jlsmith/lex-corp.html

**Google Is Your Friend, GIYF** is a term sometimes used in chat and forums to let the person asking the question know the answer could have been found by using the Google.

[https://www.computerhope.com/jargon/g/giyf.htm](https://www.computerhope.com/jargon/g/giyf.htm)
Corpus data sources and software tools
AntFileConverter (PDF & Word to plain text)
Corpus data sources and software tools

AntCorGen (PLOS One access and corpus download)

All Science Deserves to Be Published

*PLOS ONE* gives researchers a faster path to publishing in a high-quality peer-reviewed journal. All work that reaches rigorous technical and ethical standards is published and freely and immediately available to everyone.

Submit Your Manuscript

Join the worldwide community of researchers - from Nobel laureates to early career researchers - who choose *PLOS ONE*. 
Corpus data sources and software tools
AntCorGen (PLOS One access and corpus download)
Corpus data sources and software tools
TagAnt (Part-Of-Speech Tagger)
Corpus data sources and software tools

AntConc (Corpus analysis toolkit)

AntConc 3.3.5w (Windows) 2012

Corpus Files
- amex7.txt
- amex9.txt
- amex10.txt
- amex11.txt
- amex13.txt
- amex14a.txt
- amex14b.txt
- amex15a.txt
- amex15b.txt
- amex16.txt
- amex17.txt
- amex18.txt
- amex20.txt
- amex21.txt
- amex1.txt
- amex3.txt
- amex6.txt

Concordance
Concordance Hits: 106

Search Term: want

Advanced

L. Anthony (2016)
Corpus data sources and software tools

Corpus data sources and software tools

AntConc (Corpus analysis toolkit)

- Freeware
- Multiplatform
  - Win XP/7/8/10
  - Linux
  - Mac 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
New features in AntConc 3.5

- **Global features**
  - OS-intuitive shortcuts for Win, Mac, Linux
  - Backend database processing (for improved speed)
  - Auto-detect file encoding option

- **"KWIC" Concordancer**
  - "Show Every Nth Row" thinning

- **Concordancer Plot Tool**
  - "Hit numbers", "thinning", "Save", and "Relative Length"

- **Collocates**
  - "Best practice" statistics for collocate measures.
    - 'Log-Likelihood', 'MI + Log-Likelihood (p > 0.05)' + MI, T-Score

- **Word List Tool**
  - accepts a word list (or raw files)
Corpus data sources and software tools

AntConc tutorial

http://www.youtube.com/user/AntlabJPN/videos
Part 2a: DDL in the technical writing classroom
building a specialist corpus
Corpus building
Processing steps to building a corpus

Choose a target area

Search for pre-built corpora in target area

Investigate language features

Design your own corpus (DIY)

Decide a sampling procedure

Collect, clean, tag/annotate
Case 1: Building your own specialist corpus
Manual approach

- **Step 1:**
  - Download (or get) five or more (interesting) research papers

- **Step 2:**
  - Put all texts (e.g., .pdf/.html) into a folder called "raw_texts"

- **Step 3:**
  - Copy and paste each section (title, abstract,...) into text files (.txt) and save in separate folders called "tit", "abs", "int", "met", "res", "dis", "ack", ... under one main folder "data"
    - When saving files, use the UTF-8 character encoding option (if asked by the system)
Case 1: Building your own specialist corpus  
Semi-automatic approach

- **Step 1:**
  - Download (or get) five or more (interesting) research papers as PDF and/or Microsoft Word (docx) files

- **Step 2:**
  - Put all texts into a folder called "raw_texts"

- **Step 3:**
  - Download AntFileConverter from www.laurenceanthony.net/software/antfileconverter/

- **Step 4:**
  - Launch AntFileConverter and drop your files into the left window (or load the files via the File menu)

- **Step 5:**
  - Click the "Start" button and wait for the corpus to be built
Case 1: Building your own specialist corpus
Fully-automatic approach

- **Step 1:**
  - Download AntCorGen from www.laurenceanthony.net/software/antcorgen/

- **Step 2:**
  - Launch AntCorGen and select a folder into which your corpus will be saved

- **Step 3:**
  - Search/select your area of interest and the sections you want to create

- **Step 4:**
  - Click the "Create Corpus" button and wait for the corpus to be built
Part 2b: DDL in the technical writing classroom

pre-course target language identification
Case 2: Target language identification

Discipline-specific vocabulary

Paul Erdos

Figure 4

where \( f_1, \ldots, f_n \) is a segment of \( a^p \). We have

\[
(27) \quad \forall x \in [0,1), \quad f_1(x) + f_2(x) + \cdots + f_n(x) \geq x + \varepsilon + \frac{1}{2n}.
\]

Consequently,

\[
(28) \quad \forall x \in [0,1), \quad f_1(x) + f_2(x) + \cdots + f_n(x) \geq x + \frac{1}{2n}.
\]

Finally,

\[
(29) \quad \forall x \in [0,1), \quad f_1(x) + f_2(x) + \cdots + f_n(x) \geq x + \frac{1}{2n}.
\]

Let \( g_1, \ldots, g_m \) denote the elements represented by \( \mathcal{H}_0 \), components of \( c \). Note that

\[
(30) \quad \forall i \geq 1, \quad g_i \mid \mathcal{H} = 1, \ldots, m; \quad \sum_{i=1}^{m} f_i(x) \leq K(x) \leq (32) + 40 \varepsilon \leq K(32x + 8 \varepsilon).
\]

Therefore, at most one of the elements \( g_1, \ldots, g_m \) has length at least

\[
(31) \quad \frac{1}{32} - 2 \varepsilon \leq K(32x + 8 \varepsilon) \leq K(32x + 7 \varepsilon).
\]

This contradicts (27). Thus \( x \) is not isolated in \( [a,b] \). By Lemma 7.3 this means that

\[
(32) \quad \forall x \in [0,1), \quad f_1(x) + f_2(x) + \cdots + f_n(x) \geq x + \frac{1}{2n}.
\]

Now assume that we have already found components \( f_1, \ldots, f_{n-1} \) of \( c \). Let \( f_n \) be an \( \mathcal{H}_0 \) component that is not connected to \( f_1, \ldots, f_{n-1} \) respectively. The inductive step is similar to the above considerations. For definiteness, we assume that \( f_n \) is an \( \mathcal{H}_0 \) component. Then \( f_n \) is an \( \mathcal{H}_0 \) component by the choice of \( W \). We denote by \( \ell \) a path of length \( \leq 1 \) placed by an element of \( \mathcal{H}_0 \) that connects \( f_n \) to \( f_i \) (Figure 4). If \( f_n \) is isolated in the cycle \((f_1, f_2, \ldots, f_n)\), we denote by \( \ell \) the maximal initial path of the segment \([0,1, \ldots, n]\) of \( p_n \) such that \( f_i \) contains \( f_{n+1} \) and all \( \mathcal{H}_0 \) components of \( f_n \) are isolated in \( c \). As above, we can find a path \( \ell \) such that \( f_n \equiv f_{n+1} \). Since \( f_{n+1} \leq \frac{1}{2} + \varepsilon \), and so \( \mathcal{H}_0 \) component of \( f_n \) is connected to an \( \mathcal{H}_0 \) component of \( f_i \). The inequalities (27) remain valid and we arrive at a contradiction in the same way. Thus \( f_n \) is not isolated in \( c \), i.e., it is connected to a component \( f_{n+1} \) of the segment \([0,1, \ldots, n]\) of \( c \). This completes the inductive step.
Case 2: Target language identification
Discipline-specific vocabulary

Word Ranks in a Research Paper

Frequency

- the
- in
- a
- analysis
- increase
- meteorological
- poverty
Case 2: Target language identification
Discipline-specific vocabulary

- **Step 1:**
  - Load your target corpus into AntConc

- **Step 2:**
  - Generate a list of all the word frequencies in the corpus
    - What words have the highest frequencies? Why?

- **Step 3:**
  - Generate a list of keywords for the corpus by comparing the target corpus against a reference corpus of general English
    - **Hint:** add a reference corpus list (e.g. AmE06) in the settings page
    - What words have the highest keyness values? Why?

- **Step 4:**
  - Investigate how high ranked words and keywords are used in writing using the Concordance Tool
### Case 2: Target language identification

#### Discipline-specific keywords

<table>
<thead>
<tr>
<th>Rank</th>
<th>Physics</th>
<th>Math</th>
<th>Biology</th>
<th>Chemistry</th>
<th>Comp. Sci.</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
Part 2c: DDL in the technical writing classroom

in-course uses of corpora and corpus tools
Case 3: Discipline-specific writing
The CELESE Program at Waseda University
Case 3: Discipline-specific writing
The CELESE Program at Waseda University

- **Schools/Departments:**
  - 3 schools; 17 departments
  - ≈10,000 students (6,000 undergrad; 4000 grad)
  - ≈75% of students proceed to graduate school

- **English Faculty:**
  - 9 full-time faculty, 65 part-time faculty

- **English Courses:**
  - 17 courses
  - 500 sections
    - (22 sections of Technical Writing)
Case 3: Discipline-specific writing
3rd year/4th year technical writing

Course goals
- Understanding the research paper...
  - audience, purpose
  - organization
  - flow, style, presentation
- Writing a research paper
  - home-based project
  - based on lab experiment
  - thesis research
Case 3: Discipline-specific writing
3rd year/4th year technical writing: Course textbook

- **Section 1: Basic Principles**
  - Introduction to science and engineering
  - Writing principles in science and engineering (APOFSP - Swales, 2004)
  - Introduction to text analysis (biography writing)

- **Section 2: Applications in Writing**
  - Titles
  - Introductions
  - Materials and Methods
  - Results and Discussion
  - Abstracts
Case 3: Discipline-specific writing

The example of biographies

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 new-view synthesis.

Andrew Fitzgibbon studied mathematics and computer science at the University College Cork and received the PhD degree from Edinburgh University in 1997. Until June 2000, he held a Royal Society University Research Fellowship at Oxford University’s Department of Engineering Science. Since then, he has joined Microsoft Cambridge Research, Cambridge, United Kingdom. His research interests are in the intersection of computer vision and computer graphics, with research in image segmentation, region segmentation, and novel views 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 Marr 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 Boujou, released by 2D3, 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.
Case 3: Discipline-specific writing
The example of biographies

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.
Taro Suzuki received the high school diploma from Sakura High School, Chiba, Japan, in 2010. He is currently a student in the Department of Applied Mathematics at Waseda University, Tokyo, Japan. In addition, he takes the class of Technical Writing 2, which trains students’ in the skill of writing technical English. His research interests include numerical modeling, numerical calculation with computers, and theory of computation.
Case 4: STEM writing
Overview of titles in research papers
Case 4: STEM writing
Overview of titles in research papers
Features of titles

- **Audience**: students, academic scholars, and working professionals
  - scientists/engineers who might not be experts
- **Purpose**:
  - (1) to inform readers about the main points,
  - (2) to appeal to the audience to read the research paper
- **Organization**: phrase, sentence, hanging
- **Flow**:
  - phrase ⇨ prepositions; sentence ⇨ verbs; hanging ⇨ colons
- **Style**: the formal academic writing style
- **Presentation**:
  - Capitalization of only the first letter of the title
  - Capitalization of the First Letters of Nouns, Adj, Adv, and Sub-Conj
  - CAPITALIZE ALL LETTERS IN ALL WORDS
Case 4: STEM writing
Overview of titles in research papers

■ Phrase Titles
  ■ "Measurement of patient outcome in arthritis."

■ Sentence Titles
  ■ "A point mutation of Tyr-759 in interleukin 6 family cytokine receptor subunit gp130 causes autoimmune arthritis."

■ Hanging Titles
  ■ "Traumatic arthritis of the hip after dislocation and acetabular fractures: Treatment by mold arthroplasty."
Case 4: STEM writing
Overview of titles in research papers

Phrase Title Structure

<table>
<thead>
<tr>
<th>Research Topic</th>
<th>Application</th>
<th>Method</th>
<th>Scope</th>
</tr>
</thead>
<tbody>
<tr>
<td>for/with/to</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>by/using/with/through</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>in/on/at/of/over</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Novelty/Value**

(novel/interesting/valuable features of the research)

= nouns / -ing verb forms / past participles / adjectives
  that modify the topic, applications, and methods
Case 4: STEM writing
Overview of titles in research papers

**Sentence Title Structure**

- [Subject]
  - Research Topic
- [Verb/Object]
  - Novelty/Value
- [Modifier]
  - Application, Method, Scope

**Hanging Title Structure**

- [Object of Research]
  - Topic/Name/Approach
- [Connector]
  - :
- [Nature of Research]
  - Type of Research, Description Application/Method/Scope
Case 4: STEM writing
Analyzing titles in research papers

- **Part 1:** Create a corpus of research paper titles in your field
- **Part 2:** Using *AntConc*, quickly scan through your titles
  - What structure is normally used: phrase titles, sentence titles, or hanging titles?
- **Part 3:** Using the *AntConc* Concordance Tool, investigate how researchers explain the important points of their research in their titles
  - What prepositions are used to signal the scope, method, and applications?
  - How is the novelty/value of the research signalled?
- **Part 4:** Write a title for own research based on the analysis
  - What is topic, scope, method, application (if any)?
  - How will you signal the novelty/value?
Case 4: STEM writing
Overview of materials/methods in research papers
Case 4: STEM writing
Overview of materials/methods in research papers

Materials

Discounts

Dusties

1 cup whole wheat flour
1/3 cup shortening
Pinch of salt
4 tablespoons water
Your choice of favorite preserves
Powdered sugar

Methods

Cut shortening into flour until the texture of mix is crumbly. Add 4 tablespoons water and work into a dough. On floured surface roll out dough to 1/8 inch thickness. Cut 4 inch circles, the recipe should yield eight circles. Put 3/4 tablespoon of preserves in the center of each circle then fold circle in half. Crimp edges of half circle with tines of fork to seal. Bake at 425 degrees for 10-12 minutes. Dust with powdered sugar after baking.
Case 4: STEM writing
Overview of materials/methods in research papers

Materials and Methods
Case 4: STEM writing
Overview of materials/methods in research papers

- **Audience**
  - Lab report: broad audience of *novice students* (and evaluating teachers)
  - Research paper: narrow audience of *experts in the field*

- **Purpose**
  - Lab report: instruction (equipment, procedure, safety measures)
  - Research paper: summary (novel methods, special equipment, ...)

- **Organization / Flow**
  - Materials → Methods (procedures)
  - Avoid linking phrases (e.g., "first," "second," "third," "next," "then."

- **Style**
  - *past tense* verbs, *passive voice* (except in figs/tables and dimensions)
  - spell out numbers one to nine (except in units of time and measures)
  - write out numbers at the beginning of the sentence

- **Presentation (varies depending on journal)**
  - name (e.g., *Materials and Methods*, *Experimental Section*)
  - position of the section (e.g. after the *Intro.*, after the *Results*)
Case 4: STEM writing
Overview of materials/methods in research papers

- **Style - Past, passive verbs**
  - **BEFORE (Not good)**
    - "First, we assemble a lithium-ion cell with a nominal capacity of 650 mAh."
    - "First, assemble a lithium-ion cell with a nominal capacity of 650 mAh."
  - **AFTER (Good)**
    - "A lithium-ion cell with a nominal capacity of 650 mAh was assembled.“

### Common Verbs used in past, passive

<table>
<thead>
<tr>
<th>Verb Used</th>
<th>Verb Used</th>
<th>Verb Used</th>
<th>Verb Used</th>
</tr>
</thead>
<tbody>
<tr>
<td>was added</td>
<td>was determined</td>
<td>was measured</td>
<td>was set</td>
</tr>
<tr>
<td>was applied</td>
<td>was employed</td>
<td>was obtained</td>
<td>was tested</td>
</tr>
<tr>
<td>was calculated</td>
<td>was fixed</td>
<td>was performed</td>
<td>was supplied</td>
</tr>
<tr>
<td>was carried out</td>
<td>was heated</td>
<td>was placed</td>
<td>was coated</td>
</tr>
<tr>
<td>was conducted</td>
<td>was made</td>
<td>was prepared</td>
<td>was collected</td>
</tr>
</tbody>
</table>
Case 4: STEM writing
Overview of materials/methods in research papers

- **Style - Present, active voice verbs**
  - **BEFORE (Not good)**
    "The diameters of their bottom parts were typically 100-200 nm, and the lengths were varied from 500 nm to 1 µm."
  - **AFTER (Good)**
    "The diameters of their bottom parts are typically 100-200 nm, and the lengths vary from 500 nm to 1 µm."

**Common Verbs used in present, active**

<table>
<thead>
<tr>
<th>consists</th>
<th>gives</th>
<th>means</th>
<th>shows</th>
</tr>
</thead>
<tbody>
<tr>
<td>contains</td>
<td>includes</td>
<td>occurs</td>
<td>suggests</td>
</tr>
<tr>
<td>decreases</td>
<td>increases</td>
<td>provides</td>
<td>summarizes</td>
</tr>
<tr>
<td>exhibits</td>
<td>indicates</td>
<td>reaches</td>
<td>supports</td>
</tr>
<tr>
<td>falls</td>
<td>makes</td>
<td>refers</td>
<td>uses</td>
</tr>
</tbody>
</table>
Case 4: STEM writing
Overview of materials/methods in research papers

- **Style - Present, active voice verbs**
  - BEFORE (Not good)
    - "Figure 2 showed that the strength of the material decreased over time."
  - "Figure 2 was showed that the strength of the material decreased over time."
  - AFTER (Good)
    - "Figure 2 shows that the strength of the material decreased over time."
    - "Figure 2 shows that the strength of the materials decreases over time."

**Common Verbs used in present, active**

<table>
<thead>
<tr>
<th>consists</th>
<th>gives</th>
<th>means</th>
<th>shows</th>
</tr>
</thead>
<tbody>
<tr>
<td>contains</td>
<td>includes</td>
<td>occurs</td>
<td>suggests</td>
</tr>
<tr>
<td>decreases</td>
<td>increases</td>
<td>provides</td>
<td>summarizes</td>
</tr>
<tr>
<td>exhibits</td>
<td>indicates</td>
<td>reaches</td>
<td>supports</td>
</tr>
<tr>
<td>falls</td>
<td>makes</td>
<td>refers</td>
<td>uses</td>
</tr>
</tbody>
</table>
Case 4: STEM writing
Overview of materials/methods in research papers

- Overcharge reaction of lithium-ion batteries
- Takahisa Ohsakia,* Takashi Kishia, Takashi Kubokia, Norio Takami a, Nao Shimura b, Yuichi Sato b, Masahiro Sekino b, Asako Satoh b

a Corporate Research & Development Center, Toshiba Corporation, 3-4-10, Minami-Shinagawa, Shinagawa-ku, Tokyo 140-0004, Japan.
b Battery & Energy Division, Display Devices & Components Control Center, Toshiba Corporation, 3-4-10, Minami-Shinagawa, Shinagawa-ku, Tokyo 140-0004, Japan

Journal of Power Sources 146 (2005) 97–100
Fig. 1. Voltage and cell temperature profiles of a 633048-type prismatic cell at a 1 C rate overcharge.
Prismatic, hermetically sealed 633048-type lithium-ion cells with a nominal capacity of 650 mAh were assembled. The cells were 6.3mm in thickness, 30mm in width and 48mm in height, using LiCoO2 cathodes, graphitized mesophase-pitch-based carbon fiber (MCF) anodes and polyethylene separator. The graphitized MCF was prepared at Petoca Co. Ltd. The electrolyte was a 1M solution of LiPF6 in a mixture of ethylene carbonate (EC) and ethyl methyl carbonate (EMC). The overcharge test was carried out at a 1 C rate constant current using a 7.5V power supply. Current limiting or temperature trip safety devices (e.g. PTC) were not used in the experimental cells. An H-shaped glass cell was also used to analyze the gas composition evolved at the cathode and anode, respectively. The cathode (40mm × 30 mm) and anode (40mm × 30 mm) were arranged separately in the glass cell, and then overcharge was carried out at 3mAcm−2. The internal gas was collected by a micro-syringe, and analyzed by gas chromatography.

These overcharged cells were disassembled in an argon filled glove box, and the overcharged cathodes and anodes, and the separators were examined by X-ray diffraction analysis, thermal analysis and energy dispersive X-ray fluorescence spectroscopy (EDX). The thermal behaviors of the overcharged cathodes and anodes were investigated in order to understand the thermal runaway mechanism.
Case 4: STEM writing
Overview of materials/methods in research papers

- Prismatic, hermetically sealed 633048-type lithium-ion cells with a nominal capacity of 650 mAh were assembled.
- The cells were 6.3mm in thickness, 30mm in width and 48mm in height, using LiCoO2 cathodes, graphitized mesophase-pitch-based carbon fiber (MCF) anodes and polyethylene separator.
- The graphitized MCF was prepared at Petoca Co. Ltd.
- The electrolyte was a 1M solution of LiPF6 in a mixture of ethylene carbonate (EC) and ethyl methyl carbonate (EMC).
- The overcharge test was carried out at a 1 C rate constant current using a 7.5V power supply.
- Current limiting or temperature trip safety devices (e.g. PTC) were not used in the experimental cells.
- An H-shaped glass cell was also used to analyze the gas composition evolved at the cathode and anode, respectively.
- The cathode (40mm × 30 mm) and anode (40mm × 30 mm) were arranged separately in the glass cell, and then overcharge was carried out at 3 mA cm⁻².
- The internal gas was collected by a micro-syringe, and analyzed by gas chromatography.
Case 4: STEM writing
Overview of materials/methods in research papers

- A with B were assembled.
- A were 6.3mm in thickness, 30mm in width and 48mm in height, using C, D and E.
- D was prepared at F.
- G was H in I and J.
- K was carried out at L using M.
- N or O were not used in P.
- Q was also used to analyze R evolved at S and T, respectively.
- S and T were arranged separately in U, and then V was carried out at 3 mA cm\(^{-2}\).
- W was collected by X, and analyzed by Y.
Part 1: Instructions

- Follow the instructions below and investigate voice (active/passive) and tense (past/present) usage in materials and methods sections.

You will need your own corpus of materials and methods sections saved as text files.

- 1. Using the AntConc cluster tool, find the most common verbs used in the passive voice. Hint: Search for "was|were" for words clusters of two words. Write down your results.
- 2. Using the Concordance Plot tool, compare the number of times authors use "is" and "are" compared with "was" and "were." Do authors tend to write in the present or past tense?
- 3. Some authors choose to write methods mostly in the present tense. Using the Concordance Plot tool, search for "is|are" write down how many authors adopt this strategy.
- 4. How do the majority of authors use "is" and "are" in the methods section? Write three different reasons for using these verbs.
Part 2: Instructions

Follow the instructions below and investigate preposition (*at, by, during, from, in, into, on, through, to, under, via*) usage in materials and methods sections.

1. Using the Concordance tool, search for the following prepositions and write down common expression(s) in which they appear.
   - Hint: search for "at|by|in|on|under" and order by the center word (0), the first word to the right (R1), and the second word to the right (R2).
Case 4: STEM writing
Analyzing materials/methods in research papers

Part 2: Instructions

Follow the instructions below and investigate preposition (*at, by, during, from, in, into, on, through, to, under, via*) usage in materials and methods sections.

2. The preposition "to" is often combined with a verb at the start of sentences.
   - Using the Clusters tool, find the two most common verbs that can appear after "to" in this position.
   - What article typically follows the verb in these cases?
     - Hint: Set the minimum and maximum cluster size to 3, and fix the search term position to "On Left" to ensure that all the clusters start with "To". Also, remember to activate the Case option, so that only "To" is searched for.
Part 2d: DDL in the technical writing classroom

Introducing DDL into a full-scale writing program
DDL in a full-scale writing program

Things to consider

■ **Planning**
  ■ Design the writing program *first*
    ■ a "four-pillars" ESP model for writing course design
  ■ Create an environment for DDL activities
    ■ PC room with file transfer functionality
    ■ Standard room + teacher demo PC (+ student PCs)
  ■ Create (or adapt) materials to allow for DDL tasks
    ■ custom designed textbook (+ homework activities)
    ■ OR supplementary handouts (+ homework activities)

■ **Teacher training**
  ■ Familiarize teaching staff with DDL materials and methods
    ■ workshops, training materials, YouTube videos, MOOCs
  ■ Hire teaching staff with an understanding of DDL
Student A (Before taking class)

II. MATERIALS AND METHODS

To collect the information about blu-ray disc, first the Google search engine (3) was used by retrieving the term “bru-ray disc”, with 18,500,000 hits. To narrow these results down to fewer and practical ones, more detailed terms such as “history of blu-ray disc” were used and the result of these searches is organized as table 1. Except for several advertisement websites, most of the materials hit were read. As table 1 showed, the number of the hits is rather small. Predictably, however, some important terms appears frequently in the majority of the materials respectively.
Student A (After taking class)

III. MATERIALS AND METHODS

Silicon was chosen as the sample substance. Ten g smashed Sin powder was put into a metal cell. The cell was set in the X-ray diffractometer produced by Nipponbunnkou. Figure.1 shows the cross section of the X-ray diffractometer device. The $2\theta$-$\theta$ determination was adopted to measure the strength of diffraction. The voltage and current of X-ray lamp were set at 40kV and 30mV respectively. Copper or iron was chosen as the target for the X-ray lamp. The range of the angle was between 20 and 140 degree. The speed of the measurement was $4^\circ$/min ($2\theta$). Slit size was fixed in DS-1°, SS-1° and RS-1°.
### Before: Class usage of passive verbs (Methods section)

<table>
<thead>
<tr>
<th>achieved</th>
<th>considered</th>
<th>investigated</th>
</tr>
</thead>
<tbody>
<tr>
<td>analyzed</td>
<td>consulted</td>
<td>measured</td>
</tr>
<tr>
<td>applied</td>
<td>discussed</td>
<td>obtained</td>
</tr>
<tr>
<td>carried</td>
<td>divided</td>
<td>performed</td>
</tr>
<tr>
<td>collected</td>
<td>established</td>
<td>prepared</td>
</tr>
<tr>
<td>compared</td>
<td>examined</td>
<td>proposed</td>
</tr>
<tr>
<td>conducted</td>
<td>included</td>
<td>read</td>
</tr>
<tr>
<td>connected</td>
<td>introduced</td>
<td>substituted</td>
</tr>
</tbody>
</table>
### After: Class usage of passive verbs (Methods section)

<table>
<thead>
<tr>
<th>added</th>
<th>defined</th>
<th>followed</th>
<th>obtained</th>
<th>set</th>
</tr>
</thead>
<tbody>
<tr>
<td>amplified</td>
<td>described</td>
<td>generated</td>
<td>outputted</td>
<td>shown</td>
</tr>
<tr>
<td>analyzed</td>
<td>designed</td>
<td>immersed</td>
<td>passed</td>
<td>sonicated</td>
</tr>
<tr>
<td>based</td>
<td>determined</td>
<td>increased</td>
<td>performed</td>
<td>stocked</td>
</tr>
<tr>
<td>collected</td>
<td>diluted</td>
<td>incubated</td>
<td>prepared</td>
<td>stopped</td>
</tr>
<tr>
<td>compared</td>
<td>divided</td>
<td>initiated</td>
<td>published</td>
<td>substituted</td>
</tr>
<tr>
<td>conducted</td>
<td>dried</td>
<td>inserted</td>
<td>purified</td>
<td>transformed</td>
</tr>
<tr>
<td>connected</td>
<td>eluted</td>
<td>introduced</td>
<td>put</td>
<td>used</td>
</tr>
<tr>
<td>considered</td>
<td>employed</td>
<td>lysed</td>
<td>read</td>
<td>visualized</td>
</tr>
<tr>
<td>consisted</td>
<td>expanded</td>
<td>measured</td>
<td>regarded</td>
<td>washed</td>
</tr>
<tr>
<td>converted</td>
<td>expressed</td>
<td>mediated</td>
<td>removed</td>
<td></td>
</tr>
<tr>
<td>decreased</td>
<td>factorized</td>
<td>multiplied</td>
<td>separated</td>
<td></td>
</tr>
</tbody>
</table>
DDL in a full-scale writing program
Selected student feedback on using DDL

<table>
<thead>
<tr>
<th>Question (Scale 1 to 5)</th>
<th>Class One (Fri. 3) (mean of 8 students)</th>
<th>Class Two (Fri. 4) (mean of 12 students)</th>
</tr>
</thead>
<tbody>
<tr>
<td>The class was easy to understand.</td>
<td>4.9</td>
<td>4.8</td>
</tr>
<tr>
<td>The class increased my motivation.</td>
<td>4.9</td>
<td>4.6</td>
</tr>
<tr>
<td>The class improved my writing</td>
<td>4.4</td>
<td>4.6</td>
</tr>
<tr>
<td>The class was important to me.</td>
<td>4.6</td>
<td>4.7</td>
</tr>
<tr>
<td>It was a good class (overall).</td>
<td>5.0</td>
<td>4.8</td>
</tr>
</tbody>
</table>
DDL in a full-scale writing program
Selected student feedback on using DDL

- **Positive**
  - “It was really useful. I think I will use it from now.”
  - “The explanation was good. I will use it everyday.”
  - “I think it is really great.”
  - “It was good.”

- **Negative**
  - “It would be good to have a user manual.”
  - “I want to know more about how to use AntConc.”
  - “I want to know more about good search terms.”
  - “I wanted a more detailed explanation of the tool.”
  - “Making the corpus was annoying.”
Part 3: Discussion
Challenges and limitations of DDL; Future work in DDL
Challenges and limitations of DDL

- Many teachers are not 'techies'.
  - Can they confidently use and demonstrate corpus tools?
    - Remember that it is a writing classroom. The use of tools should be minimal

- Many universities do not have enough PC labs for writing classes.
  - Can the corpus approach be brought into the classroom?
    - Activities using *AntConc* can be completed for homework

- Most writing textbooks do not encourage DDL
  - How easy can traditional materials be adapted to suit DDL?
    - Minimally use corpora and corpus tools to simply confirm/reject textbook/classroom generalizations
    - Ideally, design new corpus-focus writing materials
Challenges and limitations of DDL

- **Corpora and corpus tools are more suited to bottom-up, word-level analysis of writing**
  - Collaborate with corpus tools' developers to improve current software tools to work at the discourse/rhetorical level

- **Corpus tools are often unintuitive and difficult to use**
  - Collaborate with corpus tools' developers to create more user-friendly, classroom-oriented tools

- **The effectiveness of DDL can be difficult to measure**
  - Carry out empirical research to measure the effectiveness
  - Analyze student writing to get a holistic view of change
  - Carry out student questionnaires to assess short-term and long-term update of DDL methods
Challenges and limitations of DDL

- Can you see yourself using the DDL approach in your own writing classroom?
  - If yes, what is the greatest advantage?
  - If no, why not?

- What do you think is the greatest challenge to using the DDL approach in a discipline-specific writing classroom?
  - limitations of current corpora?
  - limitations of current corpus tools?
  - (lack of) teacher experience and knowledge of computers
Future work in DDL:
Social Network Analysis and Data Mining

Future work in DDL: Social Network Analysis and Data Mining

FireAnt visualization output (exported to Gephi)

Future work in DDL: Social Network Analysis and Data Mining

Analysis of research article titles in *English for Specific Purposes*

Future work in DDL:
Cognitive Psycholinguistics (e.g. Eye-Tracking)

Detecting eye-fixations with *EyeChat*

Future work in DDL:
Cognitive Psycholinguistics (e.g. Eye-Tracking)

- **NS1 - fixated words (based on *EyeChat*)**
  - 'Going overseas for university study is an exciting prospect for many people. But while it may offer some advantages, it is probably better to stay home because of the difficulties a student inevitably encounters living and studying in a different culture.'

To what extent do you agree or disagree with this statement? Give reasons for your answer and include any relevant examples from your knowledge or experience.

Discuss with your partner about this topic.
Future work in DDL: Cognitive Psycholinguistics (e.g. Eye-Tracking)

- **NNS1 - fixated words (according to EyeChat)**
  - 'Going overseas for university study is an exciting prospect for many people. But while it may offer some advantages, it is probably better to stay home because of the difficulties a student inevitably encounters living and studying in a different culture.'

To what extent do you agree or disagree with this statement? Give reasons for your answer and include any relevant examples from your knowledge or experience.

Discuss with your partner about this topic.

EyeChat
Future work in DDL: Interactive text-enhancement with *EyeChat*

Text fixations highlighted and 'enhanced'

'Going overseas for university study is an exciting **prospect** for many people. But while it may offer some advantages, it is probably better to stay home because of the difficulties a student inevitably encounters living and studying in a...

Chat fixations highlighted and 'enhanced'

eyechat_b Have you ever been to a **museum**? I love them very much.

No, I haven't. I'd love to go
Conclusions

- **Corpora and corpus tools can be very effective in the preparation and teaching of writing**
  - They provide teachers and students with ways to analyze language empirically (scientifically)
  - They empower teachers and students to discover features of writing in their own and other specialized fields
  - They result in noticeable gains in student writing production

- **Corpora and corpus tools should not be introduced into the classroom in an ad-hoc fashion**
  - Technical issues (software/hardware) need to be addressed
  - Carefully designed student materials are needed
  - Teaching training is essential