

Solving the challenge of locating, preparing, displaying, and reusing academic reading texts for learners with ADAPS

Ralph Rose <rose@waseda.jp>

Center for English Language Education
Waseda University Faculty of
Science and Engineering
Tokyo, Japan



<https://fildpauz.github.io/adaps/>



Problem: Authentic academic reading texts



For ESP, authentic texts are desirable (Carver 1983, Gilmore 2011, Hudson 1991)

But this poses several problems

- Locating suitable texts
- Locating timely texts
- Analyzing the text
- Preparing the text for teaching or learning
- Reusing texts
- Sharing and redistributing texts

“Academic Reading” at Waseda Sci/Eng



- One-year required course for 2,000 students in 48 sections taught by 30 teachers
- Coordinated curriculum involving teacher training, semi-uniform testing, and materials development
- Practical needs for materials
 - Show linguistic structure in uniform fashion
 - Allow customization of display, dynamically
 - Be accessible to both teachers and learners

Value of text highlighting



- (Brett 1997, 1998; Chun and Plass 1996, 1997; De Ridder 2002; Hulstijn et al 1996; Lomika 1998)
- Increases incidental learning of vocabulary
- Increases text comprehension
- Increases engagement time with text

ADAPS



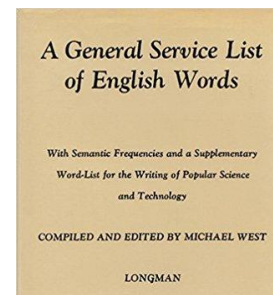
- “Academic Document Annotation and Presentation Schema”
- Defines a XML schema for annotating academic documents for various features
- Defines a JavaScript library and stylesheets for visual presentation of documents with interactive capability.

ADAPS Annotation: Vocabulary



- General Service Lists (West 1953)
 - GSL 1
 - GSL 2
- Academic Word List (Coxhead 2000)
 - Sublists 1-10
- Technical terms
- Abbreviations

“This population size, which represents the maximum population size that a particular environment can support, is called the **carrying capacity**, or **K**.”



Assume: assumed, assumes, assuming, assumption, assumptions

Authority: authoritative, authorities

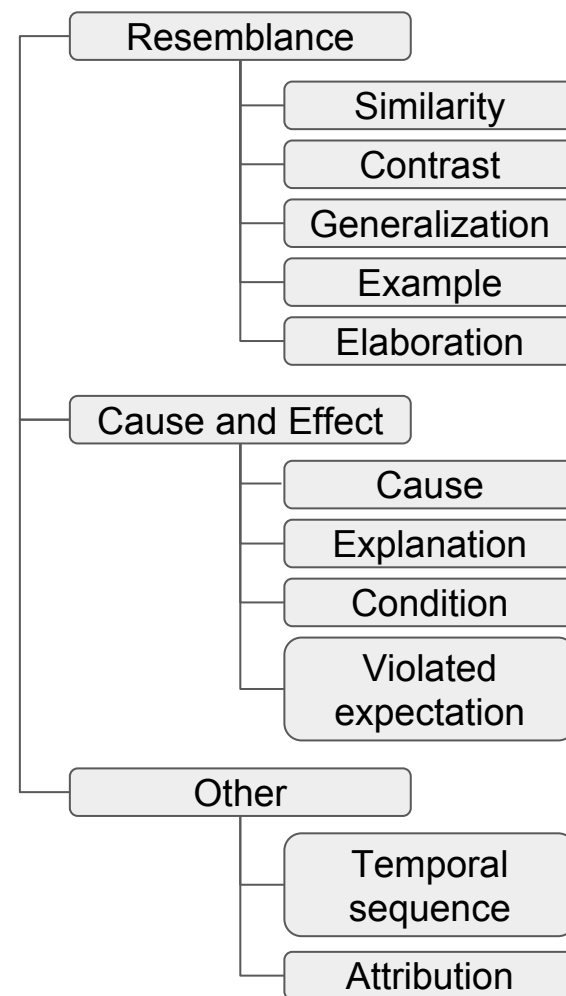
Available: availability, unavailable

Benefit: beneficial, beneficiary, beneficiaries, benefited, benefiting, benefits

ADAPS Annotation: Logical connectors



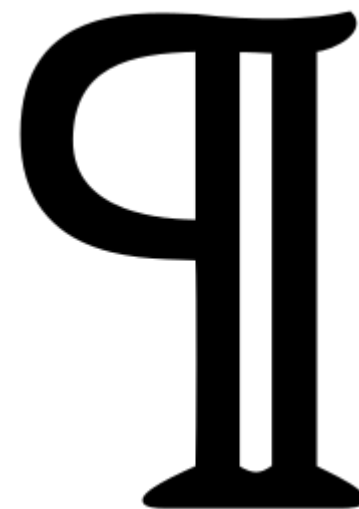
- Annotate constructions of ***x c y*** where [***x***, ***y***] are standalone clauses and [***c***] is a logical connector that defines a logical relationship between the clauses.
- Use schema adapted from Hobbs (1985) and Wolf and Gibson (2005).



ADAPS Annotation: Paragraph structure



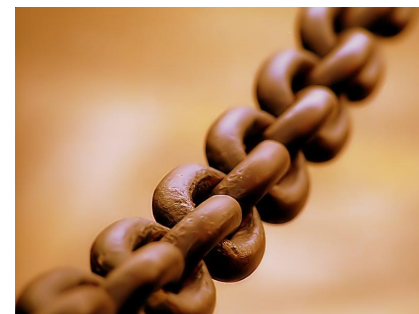
- Use classic analysis of paragraph structure: topic, supporting, and concluding sentences.
- Annotate each sentence in a paragraph according to its role.



ADAPS Annotation: Anaphoric elements



- Locate anaphoric links within documents
 - Pronominal expressions
 - Demonstrative phrases
 - Definite noun phrases
- Annotate anaphoric expressions with links to in-text referents



ADAPS Annotation: final product (XML)



```
C:\Users\Ralph\Documents\GitHub\adaps\docs\pipeline\5-detail-and-fi...
File Edit Search View Encoding Language Settings Macro Run TextFX Plugins Window ?
algorithm.xml
1 <?xml version="1.0" encoding="utf-8"?>
2 <!DOCTYPE document SYSTEM "../.../xml/adaps.dtd">
3 <document>
4   <title>Algorithm (Extract)</title>
5   <citation>Simple English Wikipedia: "Algorithm". Retrieved from
6     <a href="https://simple.wikipedia.org/wiki/Algorithm">https://simple.wikipedia.org/wiki/Algorithm</a>
7   <info type="arise">
8     <arise-info book="2" section="4" unit="3" firstpage="56"/>
9   </info>
10  <body>
11    <paragraph id="p01">
12      <sentence par-function="topic" id="p01-s01">
13        <t id="p01-s01-t01" type="GSL-1">An</t>
14        <term id="p01-s01-e01" info="p01-s01-t03:p01-s01-t19">
15          <t id="p01-s01-t02">algorithm</t>
16        </term>
17        <t id="p01-s01-t03" type="GSL-1">is</t>
18        <t id="p01-s01-t04" type="GSL-1">a</t>
19        <t id="p01-s01-t05" type="GSL-1">step</t>
20        <punc></punc>
21        <t id="p01-s01-t06" type="GSL-1">by</t>
22        <punc></punc>
23        <t id="p01-s01-t07" type="GSL-1">step</t>
24        <t id="p01-s01-t08" type="GSL-2">list</t>
25        <t id="p01-s01-t09" type="GSL-1">of</t>
26        <t id="p01-s01-t10" type="GSL-1">directions</t>
27        <t id="p01-s01-t11" type="GSL-1">that</t>
28        <t id="p01-s01-t12" type="GSL-1">need</t>
29        <t id="p01-s01-t13" type="GSL-1">to</t>
30        <t id="p01-s01-t14" type="GSL-1">be</t>
31        <t id="p01-s01-t15" type="GSL-1">followed</t>
32        <t id="p01-s01-t16" type="GSL-1">to</t>
33        <t id="p01-s01-t17" type="GSL-2">solve</t>
34        <t id="p01-s01-t18" type="GSL-1">a</t>
35        <t id="p01-s01-t19" type="GSL-1">problem</t>
36        <punc></punc>
37      </sentence>
38      <sentence par-function="supporting" id="p01-s02">
39        <anaphor id="p01-s02-a01" info="p01-s01-t05:p01-s01-t10">
40          <t id="p01-s02-t01" type="GSL-1">The</t>
41          <t id="p01-s02-t02" type="Coxhead-AWL-6,AR-AWL-6b">
            instructions</t>
          </anaphor>
        </sentence>
      </body>
    </document>
  </xml>
```

```
Ralph\Documents\GitHub\adaps\docs\pipeline\5-detail-and-fi...
arch View Encoding Language Settings Macro Run TextFX Plugins Window ?
<sentence par-function="supporting" id="p08-s06">
  <t id="p08-s06-t01">1</t>
  <punc></punc>
  <connector id="p08-s06-c01" type="condition" info="
    p08-s06-t03:p08-s06-t14,p08-s06-t15:p08-s06-t20">
    <t id="p08-s06-t02" type="GSL-1">If</t>
  </connector>
  <anaphor id="p08-s06-a01" info="p07-s01-t04:p07-s01-t12">
    <t id="p08-s06-t03" type="GSL-1">the</t>
    <t id="p08-s06-t04">stack</t>
  </anaphor>
  <t id="p08-s06-t05" type="GSL-1">of</t>
  <t id="p08-s06-t06" type="GSL-2">cards</t>
  <t id="p08-s06-t07" type="GSL-1">is</t>
  <t id="p08-s06-t08" type="GSL-2">empty</t>
  <punc></punc>
  <t id="p08-s06-t09" type="GSL-1">or</t>
  <anaphor id="p08-s06-a02" info="p07-s01-t04:p07-s01-t12">
    <t id="p08-s06-t10" type="GSL-1">it</t>
  </anaphor>
  <t id="p08-s06-t11" type="GSL-1">only</t>
  <t id="p08-s06-t12" type="GSL-1">contains</t>
  <t id="p08-s06-t13" type="GSL-1">one</t>
  <t id="p08-s06-t14" type="GSL-2">card</t>
  <punc></punc>
  <anaphor id="p08-s06-a03" info="p07-s01-t04:p07-s01-t12">
    <t id="p08-s06-t15" type="GSL-1">it</t>
  </anaphor>
  <t id="p08-s06-t16" type="GSL-1">is</t>
  <t id="p08-s06-t17" type="GSL-1">sorted</t>
  <punc></punc>
  <t id="p08-s06-t18" type="GSL-1">you</t>
  <t id="p08-s06-t19" type="GSL-1">are</t>
  <t id="p08-s06-t20" type="GSL-1">done</t>
  <punc></punc>
</sentence>
<sentence par-function="supporting" id="p08-s07">
  <t id="p08-s07-t01">2</t>
  <punc></punc>
  <t id="p08-s07-t02" type="GSL-1">Take</t>
  <anaphor id="p08-s07-a01" info="p07-s01-t04:p07-s01-t12">
    <t id="p08-s07-t03" type="GSL-1">the</t>
    <t id="p08-s07-t04">stack</t>
  </anaphor>
</sentence>
```

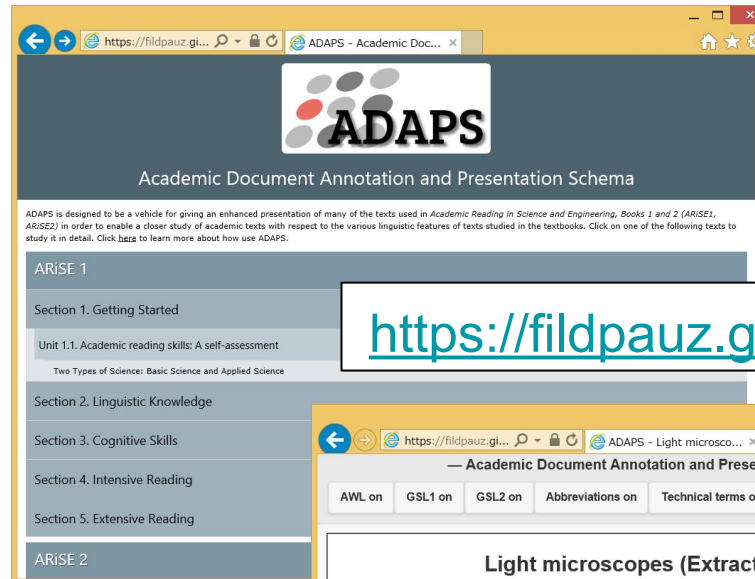
eXtensible Markup length : 51421 lines : 95 Ln : 1 Col : 1 Sel : 0 | 0 Dos¥Windows UTF-8 INS

ku length : 51421 lines : 95 Ln : 639 Col : 1 Sel : 0 | 0 Dos¥Windows UTF-8 INS

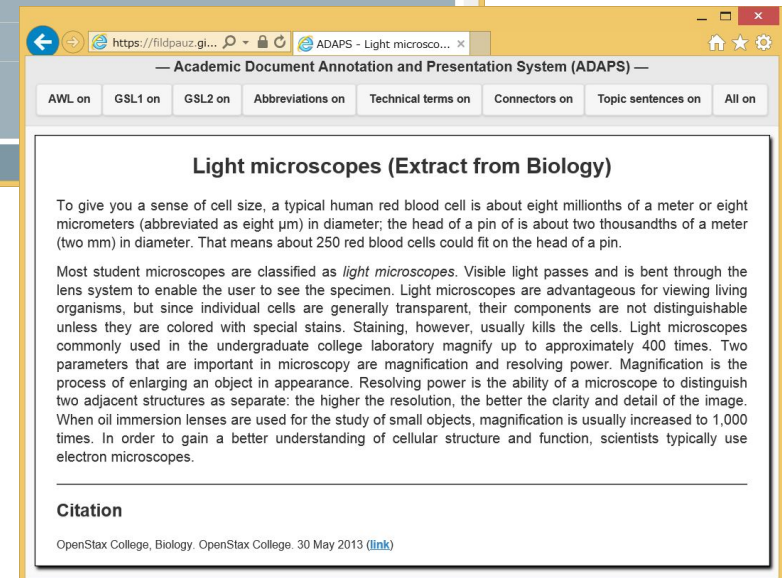
ADAPS Presentation: Table of contents



- Access is not restricted or limited.
- Users can select from a long list of documents in a variety of genres of academic writing.



<https://fildpauz.github.io/adaps/>



ADAPS Presentation: Vocabulary



- Clicking on the GSL1, GSL2, and AWL buttons will cause highlighting to appear on the vocabulary words in the text.
- Clicking on an AWL word will cause a pop-up to appear with a dictionary gloss (from LDCE online).

The screenshot shows a web browser window with the URL <https://fildpauz.gi...> and a tab titled "ADAPS - Light microscop...". The page header reads "Academic Document Annotation and Presentation System (ADAPS)". Below the header is a row of buttons: "AWL off", "GSL1 off", "GSL2 off", "Abbreviations on", "Technical terms on", "Connectors on", "Topic sentences on", and "All on". The main content area is titled "Light microscopes (Extract from Biology)". It contains two paragraphs of text with various words highlighted in green and red. A pop-up box is visible over the text, containing the text "To make it possible for someone to do something". Below the text is a "Citation" section with the text "OpenStax College, Biology. OpenStax College. 30 May 2013 ([link](#))".

Light microscopes (Extract from Biology)

To give you a sense of cell size, a typical human red blood cell is about eight millionths of a meter or eight micrometers (abbreviated as eight μm) in diameter; the head of a pin is about two thousandths of a meter (two mm) in diameter. That means about 250 red blood cells could fit on the head of a pin.

Most student microscopes use light. To make it possible for someone to do something, light passes and is bent through the lens system to enable the user to see the specimen. Light microscopes are advantageous for viewing living organisms, but since individual cells are generally transparent, their components are not distinguishable unless they are colored with special stains. Staining, however, usually kills the cells. Light microscopes commonly used in the undergraduate college laboratory magnify up to approximately 400 times. Two parameters that are important in microscopy are magnification and resolving power. Magnification is the process of enlarging an object in appearance. Resolving power is the ability of a microscope to distinguish two adjacent structures as separate: the higher the resolution, the better the clarity and detail of the image. When oil immersion lenses are used for the study of small objects, magnification is usually increased to 1,000 times. In order to gain a better understanding of cellular structure and function, scientists typically use electron microscopes.

Citation

OpenStax College, Biology. OpenStax College. 30 May 2013 ([link](#))

ADAPS Presentation: Terms & Abbreviations



- Clicking on the terms and abbreviations buttons will cause highlighting to appear on those words in the text.
- Moving the mouse cursor over a word will cause its definition to appear in a pop-up.

The screenshot shows a web browser window with the URL <https://fildpauz.gi...> and a tab titled "ADAPS - Light microscop...". The page header is "Academic Document Annotation and Presentation System (ADAPS)". Below the header is a navigation bar with buttons: "AWL on", "GSL1 on", "GSL2 on", "Abbreviations off" (highlighted in pink), "Technical terms off" (highlighted in blue), "Connectors on", "Topic sentences on", and "All on".

The main content area is titled "Light microscopes (Extract from Biology)". It contains two paragraphs of text. The first paragraph discusses cell size and uses the unit "micrometers" (abbreviated as "um"). A pop-up box labeled "Definition: micrometers" appears over the word "um". The second paragraph discusses the classification of light microscopes and their use. A pop-up box labeled "Definition: the ability of a microscope to distinguish two adjacent structures as separate" appears over the word "resolution".

At the bottom of the page, there is a "Citation" section with the text: "OpenStax College, Biology. OpenStax College. 30 May 2013 ([link](#))".

ADAPS Presentation: Logical connectors



- Clicking on the connectors button will cause logical connectors in the text to appear in red.
- Moving the mouse cursor over a connector will cause its type and the connected ideas to appear in a pop-up.

The screenshot shows the ADAPS (Academic Document Annotation and Presentation System) web interface. The browser address bar shows the URL <https://fildpauz.gi...>. The page title is "ADAPS - Light microscop...". The interface includes a navigation bar with buttons for "AWL on", "GSL1 on", "GSL2 on", "Abbreviations on", "Technical terms on", "Connectors off", "Topic sentences on", and "All on". The main content area displays a text extract titled "Light microscopes (Extract from Biology)". The text contains several logical connectors highlighted in red: "but", "since", "however", and "When". A pop-up box is visible over the text, displaying the type of the selected connector and the connected ideas. The pop-up box contains the following information:

- TYPE: Violated Expectation
- IDEA 1: their components are not distinguishable unless they are colored with special stains
- IDEA 2: Staining usually kills the cells

The text in the main content area is as follows:

To give you a sense of cell size, a typical human red blood cell is about eight millionths of a meter or eight micrometers (abbreviated as eight μm) in diameter; the head of a pin of is about two thousandths of a meter (two mm) in diameter. That means about 250 red blood cells

Most student microscopes are classified as *light microscopes* lens system to enable the user to see the specimen. Light organisms, **but since** individual cells are generally transparent unless they are colored with special stains. Staining, **however**, usually kills the cells. Light microscopes commonly used in the undergraduate college laboratory magnify up to approximately 400 times. Two parameters that are important in microscopy are magnification and resolving power. Magnification is the process of enlarging an object in appearance. Resolving power is the ability of a microscope to distinguish two adjacent structures as separate: the higher the resolution, the better the clarity and detail of the image. **When** oil immersion lenses are used for the study of small objects, magnification is usually increased to 1,000 times. In order to gain a better understanding of cellular structure and function, scientists typically use electron microscopes.

Citation

OpenStax College, Biology. OpenStax College. 30 May 2013 ([link](#))

ADAPS Presentation: Topic sentences



- Clicking on the topic sentences button will cause underlining to appear on topic sentences of paragraphs.

The screenshot shows a web browser window with the URL <https://fildpauz.gi...> and a tab titled "ADAPS - Light microscop...". The page header is "— Academic Document Annotation and Presentation System (ADAPS) —". Below the header is a row of buttons: "AWL on", "GSL1 on", "GSL2 on", "Abbreviations on", "Technical terms on", "Connectors on", "Topic sentences off", and "All on". The main content area is titled "Light microscopes (Extract from Biology)". It contains two paragraphs of text. The first paragraph is underlined: "To give you a sense of cell size, a typical human red blood cell is about eight millionths of a meter or eight micrometers (abbreviated as eight μm) in diameter; the head of a pin of is about two thousandths of a meter (two mm) in diameter. That means about 250 red blood cells could fit on the head of a pin." The second paragraph is not underlined: "Most student microscopes are classified as *light microscopes*. Visible light passes and is bent through the lens system to enable the user to see the specimen. Light microscopes are advantageous for viewing living organisms, but since individual cells are generally transparent, their components are not distinguishable unless they are colored with special stains. Staining, however, usually kills the cells. Light microscopes commonly used in the undergraduate college laboratory magnify up to approximately 400 times. Two parameters that are important in microscopy are magnification and resolving power. Magnification is the process of enlarging an object in appearance. Resolving power is the ability of a microscope to distinguish two adjacent structures as separate: the higher the resolution, the better the clarity and detail of the image. When oil immersion lenses are used for the study of small objects, magnification is usually increased to 1,000 times. In order to gain a better understanding of cellular structure and function, scientists typically use electron microscopes." Below the text is a "Citation" section with the text: "OpenStax College, Biology. OpenStax College. 30 May 2013 ([link](#))".

Choice of texts



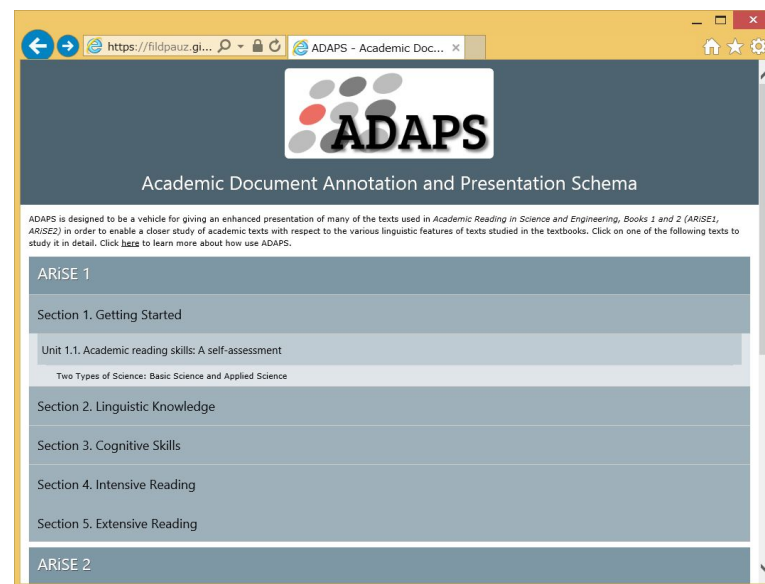
- Texts were initially chosen to support the “Academic Reading” course
 - Hence, all texts are from the textbooks used to teach the course: *Academic Reading in Science and Engineering, Book 1 and 2* (ARiSE 1, ARiSE 2)
- Other texts chosen from various genres: science news, textbooks, research articles, Wikipedia
- All texts are from open license sources



Current status



- 96 academic texts
- 442 paragraphs
- 32,670 word tokens
 - 22,804 GSL1 word tokens
 - 1,722 GSL2 word tokens
 - 6,146 AWL word tokens
- 140 technical terms
- 44 abbreviations
- 700 logical connectors
- 1,204 anaphors



<https://fildpauz.github.io/adaps/>



Summary of challenges

- Locating texts
 - Preselected texts for pedagogical purposes
- Preparing texts
 - Linguistic analysis and annotation completed
- Displaying texts
 - Flexible, dynamic, customizable presentation
- Reusing texts
 - Open license allows for unlimited reproduction (with source attribution)



Problems and limitations

- Vocabulary sense mismatches possible
 - Sense in text $\neq$ sense in vocabulary list
 - Gloss $\neq$ sense in text
- Error-checking
- Not fully mobile-compatible



Future plans

- Error-checking (second expert)
- Enable highlighting of anaphoric elements
- Improve compatibility with handheld devices
- Add more texts
- Creates indexes
 - Lists of logical connectors of a specific type
 - Examples of atypical paragraph structure
- Effectiveness research

References



Brett, P. (1997) A comparative study of the effects of the use of multimedia on listening comprehension, *System*, 25 (1): 39-53.

Brett, P. (1998) Using multimedia: A descriptive investigation of incidental language learning, *Computer Assisted Language Learning*, 11 (2): 179-200.

Carver, D. (1983), Some propositions about ESP. *The ESP Journal*, 2(2): 131-137.

Chun, D.M. and Plass, J.L. (1996), Effects of Multimedia Annotations on Vocabulary Acquisition. *The Modern Language Journal*, 80: 183–198.

Chun, D.M., and Plass, J.L. (1997). Research on text comprehension in multimedia environments. *Language learning and technology*, 1(1), 60-81.

Coxhead, A. 2000. A New Academic Word List. *TESOL Quarterly*, 34 (2): 213-238.

De Ridder, I. (2002). Visible or Invisible Links: Does the Highlighting of Hyperlinks Affect Incidental Vocabulary Learning, Text Comprehension, and the Reading Process?. *Language Learning & Technology*. 6 (1), p. 123.

Gilmore, A. (2011), "I Prefer Not Text": Developing Japanese Learners' Communicative Competence with Authentic Materials. *Language Learning*, 61(3): 786–819.

Hobbs, J.R. (1985), On the coherence and structure of discourse. Technical Report 85-37, Center for the Study of Language and Information (CSLI), Stanford, CA.

Hudson, T. (1991). A Content Comprehension Approach to Reading English for Science and Technology. *TESOL Quarterly*, 25(1), 77-104.

Hulstijn, J.H., Hollander, M. and Greidanus, T. (1996), Incidental Vocabulary Learning by Advanced Foreign Language Students: The Influence of Marginal Glosses, Dictionary Use, and Reoccurrence of Unknown Words. *The Modern Language Journal*, 80: 327–339.

LDCE online: Longman Dictionary of Contemporary English online API. <http://developer.pearson.com/apis/dictionaries>

Lomika, L.L. (1998). "To Gloss or Not To Gloss": An Investigation of Reading Comprehension Online. *Language Learning & Technology*. 1 (2), pp. 41-50.

West, M. 1953. *A General Service List of English Words*. London: Longman, Green and Co.

Wolf and Gibson (2005) "Representing discourse coherence: A corpus-based analysis". *Computational Linguistics*, 31(2): 249-287.