Writing in tables and lists: exploring multimodal undergraduate writing through keyword searches
Writing in tables and lists:
Exploring multimodal undergraduate writing through keyword searches

Maria Leedham
m.e.leedham@open.ac.uk

ICAME 2011
1. Research questions and the corpora

2. Findings
   2.1 Overall findings from keyword analysis
   2.2 Visuals and lists
   2.2 Comparison of text pairs in Biology and Economics

3. Conclusions
Why study UK undergraduate writing?

- UG assessed writing is a ‘high-stakes’ genre which has been under-researched

- UG writing is challenging due to the recent increase in multi-disciplinary degrees…

  *But…* academic writing varies between disciplines (e.g. Hewings, 1999; Hyland, 2008)

- Increase in new genres in UK assignments e.g. reflective blogs, website evaluations or press releases (Leedham, 2009; Nesi & Gardner, 2006).

- A major strategic aim of assignment-writing is to display disciplinary knowledge in an appropriate form
Why study Chinese students’ writing?

• The ‘largest single overseas student group in the UK’ (British Council, 2010)
  Over 85,000 Chinese students in UK in 2009

• BUT… most studies of Chinese students’ writing have been carried out on learner corpora (e.g. Chuang and Nesi, 2006; Mayor et al., 2007) or postgraduate theses (e.g. Hyland, 2008)

• This study uses authentic undergraduate assignments

Research Questions

1. What are some of the differences between Chinese and British undergraduate students' assessed UG writing?

2. How do these vary across the years of undergraduate study?

3. How significant is the discipline of study?
The Corpora

British Academic Written English (BAWE)

- 6,506,995 words
- 2,896 texts
- 2,761 assignments
- 1,039 contributors
- 30+ disciplines
- 13 genre families
- 4 levels of study
- Variety of L1s
- All proficient writing

The corpora for this study

- Extracted L1 English and L1 Chinese texts from BAWE
- Reduced to UG texts only
- Selected 5 disciplines
- Added extra L1 Chinese texts from other sources
- Resulting in: 104 texts from Chinese students
- 295 texts from British students
- Insights from lecturer interviews

ESRC project number
RES-000-23-0800
## The corpora

<table>
<thead>
<tr>
<th>Discipline</th>
<th>L1 Chinese corpus</th>
<th>L1 English corpus</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>No. of assignments</td>
<td>No. of words</td>
</tr>
<tr>
<td>Biological Science</td>
<td>18</td>
<td>33,633</td>
</tr>
<tr>
<td>Business</td>
<td>20</td>
<td>33,303</td>
</tr>
<tr>
<td>Economics</td>
<td>20</td>
<td>38,086</td>
</tr>
<tr>
<td>Engineering</td>
<td>20</td>
<td>35,627</td>
</tr>
<tr>
<td>Food Science</td>
<td>26</td>
<td>30,267</td>
</tr>
<tr>
<td><strong>Totals</strong></td>
<td><strong>104</strong></td>
<td><strong>170,916</strong></td>
</tr>
</tbody>
</table>
Outline

1. Research questions and the corpora

2. Findings
   2.1 Overall findings from keyword analysis
   2.2 Visuals and lists
   2.2 Comparison of text pairs in Biology and Economics

3. Conclusions
Keyness

- ‘A word which is *positively* key occurs *more* often than would be expected by chance in comparison with the reference corpus.’
  Scott, WordSmith Help files, 2010

- WordSmith Tools v.5 (Scott, 2010)

- Used log likelihood statistic, $p=.000001$

- Extracted key words and key n-grams (2-5 words)

- Compared keywords in Chi-Engineering with Eng-Engineering, and in each corpus with all-UG-BAWE, etc

- Checked to ensure each keyword or key n-gram occurs in writing from at least 5 texts and 3 students
<table>
<thead>
<tr>
<th>Keywords</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Key in Chi-Economics</strong></td>
</tr>
<tr>
<td>rate, model, output, formula, level, growth, curve, income, government, supply, students, population, dividends, per, reserves, consumption, T, aggregate, tax, Dutch, quantity, stock, portfolio, assets, inefficiency, competitive, capm, generation, repurchases, qtmark, asset, refer, cash, disposable, progress, deficit, income</td>
</tr>
<tr>
<td><strong>Key in Eng-Economics</strong></td>
</tr>
<tr>
<td>in, that, as market, however, if, policy, economy, firm, therefore, firms, had, competition, costs, hence, under, since, U.S., significant, period, shown, war, international, lower, world, did, Britain, markets, impact, profits, transport, Bertrand, railways, crises, states, you, cournot, wages, question, extent, stabilisation, British, vertical, shirking, credibility, IMF, governments</td>
</tr>
<tr>
<td><strong>Key in both</strong></td>
</tr>
<tr>
<td>price, demand, monopoly, we, than, capital, increase, higher, exchange, inflation, labour, economic, unemployment, prices, countries, money, production, cost, investment, interest, firm, foreign, crisis, trade, wage, long, marginal, F, country, Y, run, elasticity, domestic, variables, goods, exam, equilibrium, expectations, rates, short, consumers, monetary, surplus, policies, consumer, efficiency, spending, scale, fiscal, productivity, Phillips, slope, bank, central, monopolist, saving, relative</td>
</tr>
</tbody>
</table>

- first person pronouns (we, I)
- connectors
- references to tables and figures
- use of numbers in lists (denoted in WS by # in some disciplines)
First person pronouns

<table>
<thead>
<tr>
<th>per 10,000 words</th>
<th>Chi Biol</th>
<th>Eng Biol</th>
<th>Chi Engin</th>
<th>Eng Engin</th>
<th>Chi Econ</th>
<th>Eng Econ</th>
</tr>
</thead>
<tbody>
<tr>
<td>I</td>
<td>1</td>
<td>1</td>
<td>3</td>
<td>10****</td>
<td>9</td>
<td>14*</td>
</tr>
<tr>
<td>we</td>
<td>6</td>
<td>7</td>
<td>17</td>
<td>15</td>
<td>29</td>
<td>23</td>
</tr>
<tr>
<td>Total</td>
<td>7</td>
<td>8</td>
<td>20</td>
<td>25</td>
<td>38</td>
<td>37</td>
</tr>
</tbody>
</table>

* p<.05;  ** p<.01
** * p<.001;  **** p<.0001

Use of I in reflective writing in Engineering

(1) … I don't think this is what a professional engineer is, although I do think that a professional engineer must work in this ‘professional manner’ … (0354f).

(2) With hindsight I would have called everybody an hour before the meeting to make sure they were coming as I eventually had to later on in the project (0342c).
Connectors

<table>
<thead>
<tr>
<th></th>
<th>Chi-Economics</th>
<th>(raw)</th>
<th>Eng-Economics</th>
<th>(raw)</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>however</em></td>
<td>14</td>
<td>(55)</td>
<td>34****</td>
<td>(178)</td>
</tr>
<tr>
<td><em>therefore</em></td>
<td>17</td>
<td>(65)</td>
<td>23*</td>
<td>(122)</td>
</tr>
<tr>
<td><em>hence</em></td>
<td>8</td>
<td>(32)</td>
<td>16***</td>
<td>(86)</td>
</tr>
<tr>
<td><em>thus</em></td>
<td>12****</td>
<td>(44)</td>
<td>4</td>
<td>(20)</td>
</tr>
<tr>
<td><em>in contrast</em></td>
<td>2**</td>
<td>(7)</td>
<td>0</td>
<td>(1)</td>
</tr>
</tbody>
</table>

* p<.05; ** p<.01; *** p<.001; **** p<.0001

Preferred connectors… and difference in position

(3) …about the prices of other firms. However, monopolies can charge what they like. (6008q, L1 chinese)

(4) The implication of this however, is that price discrimination, which is possible through monopoly… (0399a L1 English)
Outline

1. Research questions and the corpora

2. Findings
   2.1 Overall findings from keyword analysis
   2.2 **Visuals and lists**
   2.2 Comparison of text pairs in Biology and Economics

3. Conclusions
## Keywords

<table>
<thead>
<tr>
<th>Key in Chi-Economics</th>
<th>rate, model, output, formula, level, growth, curve, income, government, supply, students, population, dividends, per, reserves, consumption, T, aggregate, tax, Dutch, quantity, stock, portfolio, assets, inefficiency, competitive, capm, generation, repurchases, qtmark, asset, refer, cash, disposable, progress, deficit, income</th>
</tr>
</thead>
<tbody>
<tr>
<td>Key in Eng-Economics</td>
<td>in, that, as market, however, if, policy, economy, firm, therefore, firms, had, competition, costs, hence, under, since, U.S., significant, period, shown, war, international, lower, world, did, Britain, markets, impact, profits, transport, Bertrand, railways, crises, states, you, cournot, wages, question, extent, stabilisation, British, vertical, shirking, credibility, IMF, governments</td>
</tr>
<tr>
<td>Key in both</td>
<td>price, demand, monopoly, we, than, capital, increase, higher, exchange, inflation, labour, economic, unemployment, prices, countries, money, production, cost, investment, interest, firm, foreign, crisis, trade, wage, long, marginal, F, country, Y, run, elasticity, domestic, variables, goods, exam, equilibrium, expectations, rates, short, consumers, monetary, surplus, policies, consumer, efficiency, spending, scale, fiscal, productivity, Phillips, slope, bank, central, monopolist, saving, relative</td>
</tr>
</tbody>
</table>

- first person pronouns (we, I)
- connectors
- references to tables and figures
- use of numbers in lists (denoted in WS by # in some disciplines)
Visuals: tables, figures, diagrams, images, photos…

Lists
- A ‘genuine’ or prototypical list contains ‘list items’, each consisting of a word or NP/VP.
- List items are separated by bullet points/hyphens/letters/numbers, or are indented.

Listlikes
- ‘paragraphs of running text carrying list-like formatting’
- ‘false lists’ or ‘list-likes’ contain larger units of text per list item (Heuboeck et al., 2005: 29).

<table>
<thead>
<tr>
<th>Carbon Content %</th>
<th>Classification</th>
<th>Uses</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.3-0.4</td>
<td>Low Carbon Steel</td>
<td>General purpose steel used for welding. Poor corrosion resistance.</td>
</tr>
<tr>
<td>0.3-0.7</td>
<td>Medium Carbon Steel</td>
<td>Used for the production of nuts, shafts, and gears. Very difficult to weld.</td>
</tr>
<tr>
<td>0.7-1.7</td>
<td>High Carbon Steel</td>
<td>Used in high stress applications, such as springs, and as cutting tools.</td>
</tr>
</tbody>
</table>

Table 1
Visually and lists

Lists
A ‘genuine’ or prototypical list contains ‘list items’, each consisting of a word or NP/VP. List items are separated by bullet points/hyphens/letters/numbers, or are indented.

Listlikes
- ‘paragraphs of running text carrying list-like formatting’
- ‘false lists’ or ‘list-likes’ contain larger units of text per list item (Heuboeck et al., 2005: 29).

Conclusions
The experiment yielded the following conclusions:
- The efficiency of a single stage centrifugal pump at high pump speed (3000 RPM) is better than it at low pump speed (2000 RPM).
- The input power with high pump speed increases faster than the one with low pump speed as discharge increases.
- The relationship between total head and discharge is not affected by pump speed, but higher pump speed provides higher total head.
## Keywords relating to visuals and lists

<table>
<thead>
<tr>
<th>L1&amp; discipline</th>
<th>Chi-Biol</th>
<th>Chi-Bus</th>
<th>Chi-Econ</th>
<th>Chi-Engin</th>
<th>Chi-Food</th>
</tr>
</thead>
<tbody>
<tr>
<td>Selected keywords</td>
<td>#</td>
<td>growth</td>
<td>#</td>
<td>curve</td>
<td>curve</td>
</tr>
<tr>
<td></td>
<td>table</td>
<td>curve</td>
<td>eq.</td>
<td>according</td>
<td>referring</td>
</tr>
<tr>
<td></td>
<td>data</td>
<td>refer</td>
<td>according</td>
<td>referring</td>
<td>statistical</td>
</tr>
<tr>
<td></td>
<td>equation</td>
<td>model</td>
<td>figure</td>
<td></td>
<td>statistical</td>
</tr>
<tr>
<td></td>
<td>figure</td>
<td>per</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>graph</td>
<td>output</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Biology
• All "Phases" are labeled on the graph. The curve of the Exponential Phase was straight, though some point lay outside this best straight line of fit (0041a).

Economics
• Actually the total loss resulting from the lower monopoly output \( (Q_M) \) is the grey triangle. The part of the grey triangle above \( P_C \) is the loss of consumer surplus (6008q).

Engineering
• According to the program and refer to the figure 4.1.1, it is easy to find… (6107d).

Food Science
• According to the 3 sets of data calculated above… (6150d).
References to visuals and lists

Biology
• All "Phases" are labeled on the graph. The curve of the Exponential Phase was straight, though some point lay outside this best straight line of fit (0041a).

Economics
• Actually the total loss resulting from the lower monopoly output (Q M) is the grey triangle. The part of the grey triangle above P C is the loss of consumer surplus (6008q).

Engineering
• According to the program and refer to the figure 4.1.1, it is easy to find… (6107d).

Food Science
• According to the 3 sets of data calculated above… (6150d).
## Visuals and lists

<table>
<thead>
<tr>
<th></th>
<th>Tables</th>
<th>Figures</th>
<th>Lists</th>
<th>Listlikes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chi-Biol</td>
<td>15****</td>
<td>25****</td>
<td>1</td>
<td>4</td>
</tr>
<tr>
<td>Eng-Biol</td>
<td>5</td>
<td>13</td>
<td>2</td>
<td>6</td>
</tr>
<tr>
<td>Chi-Econ</td>
<td>1</td>
<td>14****</td>
<td>2*</td>
<td>25****</td>
</tr>
<tr>
<td>Eng-Econ</td>
<td>0</td>
<td>12</td>
<td>1</td>
<td>3</td>
</tr>
<tr>
<td>Chi-Bus</td>
<td>2</td>
<td>2</td>
<td>6*</td>
<td>129****</td>
</tr>
<tr>
<td>Eng-Bus</td>
<td>6**</td>
<td>6**</td>
<td>3</td>
<td>23</td>
</tr>
<tr>
<td>Chi-Food</td>
<td>20*</td>
<td>6</td>
<td>5</td>
<td>82****</td>
</tr>
<tr>
<td>Eng-Food</td>
<td>14</td>
<td>6</td>
<td>4</td>
<td>18</td>
</tr>
<tr>
<td>Chi-Engin</td>
<td>10*</td>
<td>21</td>
<td>7</td>
<td>53****</td>
</tr>
<tr>
<td>Eng-Engin</td>
<td>7</td>
<td>21</td>
<td>10</td>
<td>24</td>
</tr>
</tbody>
</table>

* $p<.05$  
** $p<.01$  
**** $p<.0001$
Listlikes by yeargroup

[Bar chart showing comparisons of different yeargroups (Chi1, Chi2, Chi3, Eng1, Eng2, Eng3) in terms of listlikes per 10,000 words.]
Outline

1. Research questions and the corpora

2. Findings
   2.1 Overall findings from keyword analysis
   2.2 Visuals and lists
   2.2 Comparison of text pairs in Biology and Economics

3. Conclusions
Dinolfo et al. (2007) students ‘see’ or ‘read’ cells under a microscope and subsequently describe them.

“All-at-once’ processing of complex and often competing visual data’ compared with the ‘linear ‘one-at-a-time’ processing that occurs when we read written text line by line’ (Dinolfo et al., 2007: 401)

### Comparison of two Biology assignments

<table>
<thead>
<tr>
<th>Text feature</th>
<th>Chinese, text 0434a</th>
<th>English, text 0067b</th>
</tr>
</thead>
<tbody>
<tr>
<td>No. of pages excluding refs</td>
<td>15.5</td>
<td>9</td>
</tr>
<tr>
<td>No. of tokens (in WS)</td>
<td>3234</td>
<td>3201</td>
</tr>
<tr>
<td>No. of tables</td>
<td>2</td>
<td>0</td>
</tr>
<tr>
<td>No. of figures</td>
<td>17</td>
<td>5</td>
</tr>
<tr>
<td>Visuals as proportion of whole text</td>
<td>48% (7.5 pp)</td>
<td>22% (2pp)</td>
</tr>
<tr>
<td>Layout</td>
<td>whole page</td>
<td>2 columns</td>
</tr>
</tbody>
</table>
The role of maternal effect genes in the development of the nematode Caenorhabditis elegans

ABSTRACT

Caenorhabditis elegans (C. elegans) has been used as one of the favourite model organisms for developmental studies. Embryogenesis of C. elegans extensively relies on maternal effect genes for intrinsically asymmetric cell division and cell-cell interactions. In this review, the early embryogenesis of C. elegans, from the establishment of Anterior-Posterior polarity initiated by sperm entry to the asymmetrical cell divisions and different cell lineages induced by a variety of cell fate determinant is summarized, some of the molecular mechanisms carried out by the crucial maternally expressed cell fate determinants underlying these processes are described.

INTRODUCTION

The C. elegans and its life cycle

Caenorhabditis elegans (C. elegans) is a small (~1 mm long) free living soil nematode that has a predominantly hermaphroditic adult life. (Figure 1)

![Diagram of C. elegans](image)

Figure 1 Adult C. elegans [1] Upper diagram: differential interference contrast image of an adult C. elegans. Lower diagram: anatomical structures of adult C. elegans (schematic drawing). Middle Left scale bar: 0.1 mm

The life cycle of C. elegans contains an embryonic stage, four larval stages (L1-L4) and an adult stage. (Figure 2) Molt (apoplosis, new cuticle formation, and ecdysis) takes place at the end of each larval stage. Under certain external conditions such as starvation, a non-growing stage, dauer larva, may form through a facultative, reversible, arrest at the lethargus in the second of four cuticle molts. The life cycle is about 2 to 3 weeks. Each

The role of maternal effect genes in the development of the nematode Caenorhabditis elegans

Maternally expressed genes are essential for the correct patterning and cell-fate determination in the early Caenorhabditis elegans embryo. The PAR proteins and MEIs are responsible for initial polarisation of the zygote, skn-1 is required to specify the EMB fate, the bifunctional protein PIE-1 is required to maintain the totipotent germ cell lineage and specification of the AB lineage involves a system homologous to Notch in Drosophila. This review describes the current understanding of these molecular mechanisms in the specification of cell fates in the pregastrulation embryo.

Introduction

The potential of Caenorhabditis elegans as a model organism for the study of embryology emerged in the 1970s (Brenner, 1974). This free-living soil nematode is ideal for studying in the laboratory as it has a rapid period of embryogenesis (15 hours) and each worm has an invariant cell lineage, with exactly 959 somatic cells in the adult, which can be easily traced during development through the transparent cuticle (Sulston & Horvitz, 1977). C. elegans is a small round-worm, approximately 1 mm long, that lives for 2-3 weeks and can be fed on Escherichia coli, which allows large numbers to be conveniently raised in a Petri dish. The predominant adult form is hermaphroditic, containing both sperm and eggs and therefore reproduction is rapid, either by self-fertilization or by cross-fertilization with the rare males.

The genetics for C. elegans is advancing rapidly. It has a small genome at 8 x 10^7 bp and relatively few genes for a eukaryote – around 17,500. It was the first multicellular organism for which the genome was completely sequenced (C.elegans Sequencing Consortium 1999) and approximately 8,000 C.elegans proteins have already been matched to homologous human gene transcripts (Lai et al., 2000). Specific mutants may be produced by targeted deletion through transposon insertion or mutagenesis. Embryos may be manipulated by transformation or injection with transgenes and marker proteins such as green fluorescent protein (GFP) are easily visualized in the transparent embryos. RNA interference (RNAi) is a particularly useful technique for studying maternal effect genes by eliminating the expression of specific maternal or zygotic genes in offspring.

Reproduction

In hermaphrodite worms, fertilization occurs in the spermatheca – an organ where the sperm is stored – when mature oocytes pass from the ovary towards the vulva (Fig 1A-B). The point of sperm entry determines the posterior end of the embryo. After fertilization, a rigid, ovoid-shaped chitin eggshell called the chorion is made (Kemphues & Strome, 1997) and the long axis of this ovoid is termed the anteroposterior (a/p) axis of the embryo.
Diagrams and extended caption in 0434a (Chinese writer)

- 186 word extended caption
- different font to main text
- text wrapping of image and caption
- full sentences and in same, neutral stance as the rest of the text (e.g. use of passives, no first person pronouns, formal language)
- caption describes the process illustrated by diagram on the left
- diagram and main text are not integrated (in this case) – the extended caption functions as a freestanding text

(Cf work on multimodal texts by Kress & Van Leeuwen)
### Bulleted lists vs. connected prose in Economics

<table>
<thead>
<tr>
<th>Text feature</th>
<th>Chinese, 0155a</th>
<th>English, 0202j</th>
</tr>
</thead>
<tbody>
<tr>
<td>No. of pages excluding refs</td>
<td>6</td>
<td>6</td>
</tr>
<tr>
<td>No. of tokens (in WS)</td>
<td>3731</td>
<td>4242</td>
</tr>
<tr>
<td>No. of formulae</td>
<td>19</td>
<td>6</td>
</tr>
<tr>
<td>No. of lists</td>
<td>2</td>
<td>0</td>
</tr>
<tr>
<td>No. of listlikes</td>
<td>28</td>
<td>0</td>
</tr>
<tr>
<td>Lists and listlikes as % of whole text</td>
<td><strong>90%</strong></td>
<td><strong>0%</strong></td>
</tr>
</tbody>
</table>
Text 0155a, Chinese writer

Text 0202j, English writer
Comparisons of data across various groups

Pure Economics degree vs. non pure Economics degree: Students doing pure Economics degree scored 66.23 on average, while students doing a mixed-Economics degree scored 61.68 (very significant).

Female vs. Male: The average female students got 63.8, compared to 65.4 for male students.

UK students vs. non-UK students: On average, UK students gained 64.63 while non-UK students gained 66.38.

Number of parents who attended university: Those students whose parents never attended university achieved 64.12 on average, those with one parent attended university achieved 64.99 averagely, and those with both parents attended university achieved 65.59 on average.

To make some comments about these results, we need to break this up into sub-samples. Firstly we can break it up according to sex, as Siegfied and strand did. For males, table 2 shows that the mean score is 65.4%, which is higher than the corresponding score for females of 63.75%. This agrees with Siegried and Strand’s paper which claims males do better than females. However the standard deviation for males is lower than for females, 12.64% compared to 14.02%.

Use of lists by the Chinese student entails:

• Lower connector usage
• Fewer words needed to say the same thing
• Higher mean word length and lower mean sentence length
Interviews with lecturers

Importance of visuals

• Diagrams and formulae are ‘the spine of the essay’ (Economics)
• Including visuals helps students gain better marks as it avoids having to describe and introducing errors (Biology)
• The ‘challenge’ is ‘to marry the diagrams with the text’ (Economics)
• ‘there is no existing document out there which explains how to interpret their data’ (Biology)
• ‘the key writing skill for an economist is the ability to demonstrate in writing about a diagram an understanding of the analysis’ it represents (Economics)

Being concise

• Lecturers value writing which is ‘clear and concise’, and ‘succinct’ and dislike ‘verbosity’ (Engineering)
• ‘there’s never been a penalty for an essay that’s too short (Biology)
Conclusions

• Chinese students make greater use of visuals and lists than British students
• As all the assignments have achieved high scores, these differences are acceptable
• NB… EAP teachers do not always teach or encourage the use of visuals and lists – more training in disciplinary differences needed
• Need to read texts from the corpus to see what’s going on and what is missing/altered from the original texts
• Need disciplinary insights from lecturers (assignment-readers) and from students (assignment-writers)

Next steps…

• Establish ways of analyzing visual elements in assignments
• Talk to lecturers and students as to the range of acceptability in assignment layout/use of visuals/writing in lists
References

• Leedham, M. (2009) ‘From traditional essay to ‘Ready Steady Cook’ presentation: reasons for innovative changes in assignments’ In Active Learning in HE.