Transition to teaching with corpora

Tatyana Karpenko-Seccombe
University of Huddersfield
I teach EAP to international PhD students at the University of Huddersfield
Transitions

A transition from using corpora as a tool mainly for teaching vocabulary and collocations to teaching wider aspects academic writing through the use of corpora
“Why should L2 [second language] learners learn to use concordance information themselves? One reason is that no one can learn everything there is to know about a language from being told; at some point, the learning depends on observing how the language works and drawing one’s own conclusions from observations – this is called data-driven learning.”

(Cobb, 2010 p. 16)
Beginning: word choice

Using some corpus materials

- in the course of Academic vocabulary
- in one-to-one tutorials
<table>
<thead>
<tr>
<th>Serch</th>
<th>Frequency</th>
<th>Context</th>
<th>Overview</th>
</tr>
</thead>
<tbody>
<tr>
<td>4 CFI</td>
<td>WAC_humart</td>
<td>A: B: C</td>
<td>of such goods was controlled...</td>
</tr>
<tr>
<td>5 HEM</td>
<td>WAC_scienc</td>
<td>A: B: C</td>
<td>at school under the tricene system...</td>
</tr>
<tr>
<td>6 CNS</td>
<td>WAC_scienc</td>
<td>A: B: C</td>
<td>There is now a growing body...</td>
</tr>
<tr>
<td>7 LEE</td>
<td>WAC_scienc</td>
<td>A: B: C</td>
<td>of this approach is that there...</td>
</tr>
<tr>
<td>8 JAS</td>
<td>WAC_scienc</td>
<td>A: B: C</td>
<td>up until now there appears to...</td>
</tr>
<tr>
<td>9 AOK</td>
<td>WAC_scienc</td>
<td></td>
<td></td>
</tr>
<tr>
<td>10 CFI</td>
<td>WAC_humart</td>
<td></td>
<td></td>
</tr>
<tr>
<td>11 HEM</td>
<td>WAC_scienc</td>
<td>A: B: C</td>
<td>alwared with a goodness of attention...</td>
</tr>
<tr>
<td>12 CNS</td>
<td>WAC_scienc</td>
<td>A: B: C</td>
<td>complex network of family...</td>
</tr>
<tr>
<td>13 HEM</td>
<td>WAC_scienc</td>
<td>A: B: C</td>
<td>memory to a topic which...</td>
</tr>
<tr>
<td>14 HEM</td>
<td>WAC_scienc</td>
<td>A: B: C</td>
<td>described in the present book...</td>
</tr>
<tr>
<td>15 CMS</td>
<td>WAC_scienc</td>
<td>A: B: C</td>
<td>as a sociopolitical problem...</td>
</tr>
<tr>
<td>16 HEM</td>
<td>WAC_scienc</td>
<td>A: B: C</td>
<td>There is a...</td>
</tr>
<tr>
<td>17 HEM</td>
<td>WAC_scienc</td>
<td>A: B: C</td>
<td>and estimated risk...</td>
</tr>
<tr>
<td>18 HEM</td>
<td>WAC_scienc</td>
<td>A: B: C</td>
<td>derived agents...</td>
</tr>
<tr>
<td>19 HEM</td>
<td>WAC_scienc</td>
<td>A: B: C</td>
<td>derived agents...</td>
</tr>
<tr>
<td>20 HEM</td>
<td>WAC_scienc</td>
<td>A: B: C</td>
<td>derived agents...</td>
</tr>
</tbody>
</table>

**British National Corpus (BNC)**

- **Search**
- **Frequency**
- **Context**
- **Overview**

**AntConc Homepage**

**Latest Release**

**AntConc**

A freeware corpus analysis toolkit for concordancing and text analysis.

[AntConc Homepage] [Screenshots] [Help] [License]

**Downloads**

- Windows (3.5.8)
- Windows 64-bit (3.5.9)
- Macintosh OS X 10.6-10.12 (3.5.9)
- Linux 32-bit (3.5.8)
- Linux 64-bit (3.5.9)
This is a perfect solution to a massive problem!
acceptable alternative critical convincing credible perceived definitive definitive desirable simple suitable successful successful temporary tentative theoretical unsatisfactory unexplored unexplored workable

acceptable alternative critical convincing credible perceived definitive definitive desirable simple suitable successful successful temporary tentative theoretical unsatisfactory unexplored unexplored workable
Word choices

This is a **perfect** solution to a **massive** problem!

definitive
successful
effective
optimal
convincing
credible

serious
severe
fundamental
major
extensive
widespread
acute
‘One criticism aimed at DDL is that it ignores the “big themes” of language such as tenses’.

(Alex Boulton 2007)
Big theme: Stance
Stance ‘refers to the language features writers use to comment on their statements in order to convey their attitudes, opinions and degrees of commitment to what they say…’

(Hyland (2016).

It includes:

• **hedges and boosters** which increase or decrease the strength of what you are saying;
• **evaluative language** (or attitude markers) and
• **self-mention** (Hyland, 2016).
## Big theme: strength of claim

### Solution

<table>
<thead>
<tr>
<th>Strong</th>
<th>Cautious</th>
</tr>
</thead>
<tbody>
<tr>
<td>critical</td>
<td>acceptable</td>
</tr>
<tr>
<td>convincing</td>
<td>suitable</td>
</tr>
<tr>
<td>credible</td>
<td>temporary</td>
</tr>
<tr>
<td>definitive</td>
<td>tentative</td>
</tr>
<tr>
<td>successful</td>
<td>workable</td>
</tr>
<tr>
<td>effective</td>
<td>possible</td>
</tr>
<tr>
<td>optimal</td>
<td></td>
</tr>
<tr>
<td>perfect?</td>
<td></td>
</tr>
</tbody>
</table>
Big theme: strength of claim

Perfect solution?

Rome clearly never found a perfect solution for this problem; eventually practice. Of course there is no perfect solution. I had the easy choice of re-
migration to America seemed like the perfect solution. Initially, many of those who wa-
ked HFEA 1990. Yet the latter is not a perfect solution since there are still aspects pr-
though "best practice" seems like the perfect solution, companies find it hard to imple-
abroad. None has found the perfect solution but the debates on the
The most is hedged in academic writing by the use of *perhaps, even, among, probably, amongst, arguably, would have been* + appears in quotes
How can you explain these differences in frequencies?
Big theme: stance (boosters)

Frequencies of *definitely* in different genres

<table>
<thead>
<tr>
<th>SECTION (CLICK FOR SUB-SECTIONS) (SEE ALL SECTIONS AT ONCE)</th>
<th>FREQ</th>
<th>SIZE (M)</th>
<th>PER MIL</th>
<th>CLICK FOR CONTEXT (SEE ALL)</th>
</tr>
</thead>
<tbody>
<tr>
<td>SPOKEN</td>
<td>948</td>
<td>10.0</td>
<td>95.15</td>
<td></td>
</tr>
<tr>
<td>FICTION</td>
<td>605</td>
<td>15.9</td>
<td>38.03</td>
<td></td>
</tr>
<tr>
<td>MAGAZINE</td>
<td>391</td>
<td>7.3</td>
<td>53.84</td>
<td></td>
</tr>
<tr>
<td>NEWSPAPER</td>
<td>324</td>
<td>10.5</td>
<td>30.96</td>
<td></td>
</tr>
<tr>
<td>NON-ACAD</td>
<td>214</td>
<td>16.5</td>
<td>12.97</td>
<td></td>
</tr>
<tr>
<td>ACADEMIC</td>
<td>122</td>
<td>15.3</td>
<td>7.96</td>
<td></td>
</tr>
<tr>
<td>MISC</td>
<td>405</td>
<td>20.8</td>
<td>19.44</td>
<td></td>
</tr>
<tr>
<td>TOTAL</td>
<td>3,009</td>
<td></td>
<td></td>
<td>SEE ALL TOKENS</td>
</tr>
</tbody>
</table>
Big theme: stance (boosters)

Boosters: clearly

Why is clearly consistently used with results?
Big theme: stance (boosters)

Boosters: *clearly*

### British National Corpus (BNC)

<table>
<thead>
<tr>
<th>HELP</th>
<th>CONTEXT</th>
<th>FREQ</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>RESULTS</td>
<td>8</td>
</tr>
<tr>
<td>2</td>
<td>STUDY</td>
<td>8</td>
</tr>
<tr>
<td>3</td>
<td>ACT</td>
<td>7</td>
</tr>
<tr>
<td>4</td>
<td>CHILDREN</td>
<td>4</td>
</tr>
<tr>
<td>5</td>
<td>CASE</td>
<td>4</td>
</tr>
<tr>
<td>6</td>
<td>COURT</td>
<td>4</td>
</tr>
<tr>
<td>7</td>
<td>TEACHERS</td>
<td>4</td>
</tr>
<tr>
<td>8</td>
<td>DECISION</td>
<td>3</td>
</tr>
<tr>
<td>9</td>
<td>GOVERNMENT</td>
<td>3</td>
</tr>
<tr>
<td>10</td>
<td>INFORMATION</td>
<td>3</td>
</tr>
<tr>
<td>11</td>
<td>MATERIAL</td>
<td>3</td>
</tr>
<tr>
<td>12</td>
<td>POSITION</td>
<td>3</td>
</tr>
<tr>
<td>13</td>
<td>POWER</td>
<td>3</td>
</tr>
<tr>
<td>14</td>
<td>ROLE</td>
<td>3</td>
</tr>
<tr>
<td>15</td>
<td>SHOW</td>
<td>3</td>
</tr>
<tr>
<td>16</td>
<td>STATE</td>
<td>3</td>
</tr>
<tr>
<td>17</td>
<td>TEXT</td>
<td>3</td>
</tr>
<tr>
<td>18</td>
<td>TITLE</td>
<td>3</td>
</tr>
</tbody>
</table>
Big theme: Argumentation
Big theme: common patterns in introducing the research gap - tenses + word choice + evaluative language + prepositions

<table>
<thead>
<tr>
<th>Little research has been carried out in relation to the corresponding (i.e. consistency of performance on similar outcomes).</th>
</tr>
</thead>
<tbody>
<tr>
<td>was addressed. In the first place, since little research has actually been done to study the linguistic behaviour of children in British schools, remarkably little research has been carried out outside London, we chose to investigate a curriculum is to be based on research? Very little research has been carried out on teaching methods and the effectiveness.</td>
</tr>
<tr>
<td>of the printed word: &quot;a field where very little research had been done.&quot; (Locke 1974: 121)</td>
</tr>
<tr>
<td>was carried out at dispersed homes. However, little research has been carried out on the quantities of wood required. An extremely small quantity of imported wares. Very little research has been carried out on domestic wares and it remains far from clear how these filters operate.</td>
</tr>
<tr>
<td>round for close on two hundred years. Very little research has been conducted to discover how the barriers to effective working children adjust to parental epilepsy. Very little research has been conducted on the needs of children who have a parent in the labour market. Yet surprisingly little research has been conducted into this area. Using techniques developed during and drawn on in their teaching,</td>
</tr>
<tr>
<td>they would be best advised to consult. Very little research has been done on the literature guides and only one on their cause among adolescents and adults. As despite the Gulf's long history of oil pollution, little research has been done on the long-term impact on marine life there.</td>
</tr>
</tbody>
</table>
| . For example, Spears (1982) states that little research has been done on the cost of rehabilitation of severely eroded
Big theme: Patterns of introducing research problem

ECT-causal interaction between them. however, the problem is that, if this is true, there is nothing...
Big theme: Supporting claims, giving reasons

Stival as the basilinna, Apollodorus suggests the problem was not that she was an alien, but because the group was everybody's motivation to work. The problem seemed to occur because it could never be fixed. The philosophy of science. I have discussed the problem of demarcation in some detail because I believe observed. Because of this, Hume claims that the problem of justification is actually the problem of not seem to be any audio feedback. This can be a problem because if you set the times there is nothing. If we could conclude that Amersham does have a problem in credit control procedures because the account for the study of women, because it solves the problem of centuries of ignoring women in research focus-on-form, which can arise either because of a problem of communication or because of the occurrence of the latter circumstance is seen by many DCs as a problem because they in their opinion tend to ongoing initiatives just because they are new. This problem is known as 'strategic drift' as shown by, for example, Nietzsche, was an important topic because "the problem of the value of pity conjures up doubt, mi Guardian not behaving in this way would not be a problem for Plato, because he would be able to promote unrepresentativeness appears to be less of a problem in "Women Drug Users" because the lives of always clear what the root cause of a particular problem was. This was because of the fluctuating said meant that reform was slow-going. Maybe a main problem was never to be discovered, maybe because enforced by all Member States." This is a large problem for the environment globally both because of efforts to ignore inductive inference. For Popper the problem of induction need not be solved because there should be a question as "do we need to solve the problem of induction?" because to ask whether indu

‘PROFCORP, 95% of the tokens for problem (39 of 41) fall into a causal category whereas in STUCORP only 32% of the tokens (150 of 473) occur in a causal relation’. Flowerdew (2003: 499).
Big theme: Supporting claims, giving reasons

questions in their entirety. THIS is partly because, for reasons to be discusse
ed by feminist psychologists. THIS is partly because, like most psychologists, t
more than 5 - 10 per cent, and THIS is apparently because of the much greater eff
open about what he is doing. THIS is both because of the moral principle involv
occurring rhizobial bacteria. THIS is chiefly because the soya bean and its natu
enity in periods of transition. THIS is crucial because Marx was to go on to argue
from an expression. Sometimes THIS is desirable because the expression seems awk
the client though, at times, THIS is difficult because he has to approach someo
ality is at best secondary). THIS is disturbing because it challenges the appea
you are not a brain in a vat. THIS is false because if you were a brain in a vat
ple enjoy what they want, and THIS is important because it encourages consumptio
ost the national economy, and THIS is important because it helps people enjoy wh
ner areas was about 5 acres. THIS is important because specific mechanisms have
ct and designed environment. THIS is important because the differences relate t
op) would have been long. But THIS is incorrect because the" Although... process
ction is, Who sent the virus? THIS is instructive because it shows how an explan
triarchal relations. However, THIS is insufficient because by itself it does not
gists and to achieve success. THIS is largely because of diversions in their car
n has been much more limited. THIS is mainly because this account of the mechani
ften called a "talking cure". THIS is misleading because counselling is more tha
n's particular point of view, THIS is mistaken because it ignores the role playe
Big theme: Patterns of counter-argumentation

- Counter-argument: *One could argue that*
- Rebuttal of the counter-argument
- Restatement of own argument

One could argue that such a strong assumption of the normal situation in the sciences: one that allows experimental, as opposed to philosophical, work to proceed. True enough but, as I shall argue below, AI is not an experimental science but an engineering technique or, if
It seems wrong, therefore, to suggest, as some critics do, that the Lake District avoided the worst distresses. While enclosure came later to the Lake District than the southern parts of England, there is plenty of evidence of rural depopulation in the later years of the eighteenth century. This was because the system of land tenure was remarkably different from other areas ...
Big theme: Recurrent patterns in presenting results

In which examples authors:

1. highlight the significance of the results
2. compare their results with previous research
3. emphasise the novelty and contribution of their research
4. suggest treating results with caution
My transition

From
using corpora as a tool to teach vocabulary
to
using corpus searches to raise students awareness of a ‘bigger picture’:

• Elements of academic style
• Stance, e.g., strength of claim, using hedging and boosters
• Joining the use of tenses and the purpose of writing
• Argumentation: supporting claims, introducing the research gap or problem, giving reasons, using counter-argumentation, framing results
My transition

From

using corpus materials in a course of Academic vocabulary and occasionally in tutorials to

• teaching a course Academic writing with corpora (Beginners and Advanced)
• and using corpus materials in Literature reviews
  Writing introduction
  Writing abstracts
  Writing about methods
  Describing visual data
  Tenses and passive voice in academic writing
Thank you
Have you got any questions?

Tatyana Karpenko-Seccombe
Huddersfield University
t.karpenko-seccombe@hud.ac.uk
References, corpus tools and corpora


Chan, T. and Hsien-Chin Liou, H-Ch. Effects of Web-based Concordancing Instruction on EFL Students’ Learning of Verb – Noun Collocations


Cobb, T. Lextutor concordancer. Available at lextutor.ca/conc/eng/


Nesi, Gardner, Thompson & Wickens, BAWE (used in Lextutor) British Academic Written English corpus, developed at the Universities of Warwick, Reading and Oxford Brookes coventry.ac.uk/bawe


