First steps in DDL: What questions do students ask and what outcomes do they achieve?

Maggie Charles
University of Oxford
maggiecharles_oxford@yahoo.com
Editing your Thesis with Corpora

Context of the Course

- 90 doctoral students
- at least 1 completed thesis chapter
- Multi-disciplinary classes
- 45% natural science; 32% social science; 23% humanities
- 25 L1s: 31% Chinese; 9% Spanish; 7% German

- 2-3 courses per year (13 in total)
- one 2-hour session/week
- 6 weeks
- AntFileConverter (Anthony 2015)
- AntConc (Anthony 2016)
Course Procedure: Session 1

Session 1: Problem-solving

**Tutor:** Explanations and demonstrations (corpora, AntConc, **Concordance searches:** simple, wildcard, context, case-sensitive, multiple terms)

**Students:** Hands-on concordancing on a corpus of theses

**Homework:** Preparation of own questions

- *Look at your own writing; bring with you at least 5 problems to check in your corpus.*
Session 2: Building your corpus and answering your own queries

**Tutor:** Demonstration (AntFileConverter)

**Students:** Hands-on corpus building using own research articles for initial DIY corpus (10-15 articles)

- Searches in own corpus for own issues, problems
- **Worksheets completed, collected for checking by tutor**
<table>
<thead>
<tr>
<th>LANGUAGE ISSUE/PROBLEM</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
</tr>
</tbody>
</table>

**QUERY**

<table>
<thead>
<tr>
<th>SEARCH TERM AND TOOL USED</th>
<th>RESULT</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**IMPROVEMENT/COMMENT**

**OUTCOME**
Estimated Size of Initial RA Corpora in Words (n = 63)

- 10% under 50,000
- 43% 50,000 - 99,999
- 30% 100,000 - 149,999
- 11% 150,000 - 199,999
- 6% over 200,000
Data and Method

63 worksheets
Language Issue/Problem called here ‘Query’ (Q)
Improvement/Comment called here ‘Outcome’ (O)

206 query-outcome sequences
Mean number of queries per student: 3.3

Queries and Outcomes were
- Transcribed
- Coded
- converted to plain text
- examined using AntConc
Research Questions

Research question 1
What language issues/problems do DDL novices raise for corpus consultation when not prompted by an assigned task?

Research question 2
What outcomes do DDL novices achieve when using corpus searches to address their own language issues/problems?
Data Analysis: Query Types

Query Type

- **Verification**: check a given item
  - e.g. ‘Insights in’ or ‘into’?
- **Elicitation**: ascertain an unknown item
  - e.g. *Which verb to use with ‘style’?*

cf. ‘pattern-defining’ and ‘pattern-hunting’ (Kennedy & Miceli 2010)
‘Confirmatory’ and ‘compensatory’ (Yoon C. 2016)
Data Analysis: Language

Language Categories

- **Phrase**: query about a sequence of words as a whole
- **Word class** (preposition, noun, verb, linking adverb, adjective etc.): query about a single word or short fixed phrase (e.g. ‘in addition’)
- **Spelling**
- **Punctuation**
**Data Analysis: Outcomes**

<table>
<thead>
<tr>
<th>Category</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Satisfactory</td>
<td>• student found information to answer their query to their own satisfaction</td>
</tr>
<tr>
<td>Partial</td>
<td>• student found some information, but was not completely satisfied with it</td>
</tr>
<tr>
<td>Unsatisfactory</td>
<td>• student did not find any information to answer their query</td>
</tr>
</tbody>
</table>
Satisfactory
Q: existing procedures lack of or in selectivity [av prep s]
O: lack of selectivity [s]

Partial
Q: not clear about how to use the word mediation [av N p]
O: mediation of can be used. But I was criticised by my examiners for my use of nominalisation mediation of and still not sure what is wrong with it. [p]

Unsatisfactory
Q: can I use have had? [av Vb u]
O: have had – no hits. [u]
Satisfactory
Q: adjectives used with \textit{composition}. [ae \textit{adj s}]
O: \textit{one-corner composition}; \textit{high-distance composition}. Discovered 2 terms that I don’t know, but I should! [s]

Partial
Q: The use of preposition to go with the noun \textit{touch} [ae \textit{prep p}]
O: I can’t tell about it because the corpora is still small but I can use \textit{touch} with \textit{on/with/of/in}. I would like to know if \textit{touch onto} is acceptable or not, but it didn’t come up. [p]

Unsatisfactory
Q: alternative words for \textit{readily computable diagnostic tool} in \textit{these indices for readily computable diagnostic tool}. [ae \textit{phr u}]
O: not sure how to do this. [u]
Query Types (n = 206)

- Elicitations are low
- Yoon (2016) = 40%

WHY?

- Focus of the course is on editing
- Verifications may be easier for novice users
Outcomes by Query Type

Verification (n = 169)
- 83% satisfactory
- 12% partial
- 5% unsatisfactory

Elicitation (n = 37)
- 89% satisfactory
- 8% partial
- 3% unsatisfactory

- Little difference between outcomes for verification and elicitation queries
- Contrast Yoon’s (2016) success rates
  - confirmatory (verification) problems = 76.8%
  - compensatory (elicitation) problems = 59.2%
Are their elicitation queries more basic, easier to formulate?

Elicitation queries with a clear search word(s)

Q: Most used adjectives for language [ae adj s]
Q: correct use of articles environmentalism [ae art s]
Q: what’s the preposition used together with status? [ae prep s]

Elicitation queries for alternatives

Q: I often use it was shown that or it was indicated that. - I'd like search other expressions using corpus. it was * that? [ae Vb s]
O: Search: it was * that. Results: it was found that; it was suggested that; it was indicated that; it was observed that; it was reported that etc. I can also use the words above. [s]
How are other elicitation queries being searched?

Creative search solutions
1. Search words as probes
   Q: comparison – **how to compare**? [ae phr s]
   O: Search *more; smaller*. *A are more likely than B to do…; as behaviour has changed more than B; smaller than was anticipated. More likely. [s]*

2. **Terminology** as search term for investigating context and rephrasing options
   Q: I repeat similar sentences again and again. - Is it possible to use corpus to find a way of **rephrasing**? I focused this sentence this time. *The role of corticothalamic feedback system is sharpening receptive fields and modulate their temporal precision.* [ae phr s]
   O: Search: *receptive field* and *temporal precision* to find alternative verbs. Results: *altered receptive field* [s]
Top language categories = 82% of total; other categories = under 10 hits

Differences in the literature between categories of analysis

*BUT*

Phrase and preposition categories tend to be high (e.g. Park 2012)
<table>
<thead>
<tr>
<th>Language Category (n)</th>
<th>Satisfactory</th>
<th>Partial</th>
<th>Unsatisfactory</th>
</tr>
</thead>
<tbody>
<tr>
<td>verb (24)</td>
<td>96%</td>
<td>0%</td>
<td>4%</td>
</tr>
<tr>
<td>linking adverb (19)</td>
<td>90%</td>
<td>5%</td>
<td>5%</td>
</tr>
<tr>
<td>preposition (38)</td>
<td>84%</td>
<td>13%</td>
<td>3%</td>
</tr>
<tr>
<td>phrase (56)</td>
<td>84%</td>
<td>7%</td>
<td>9%</td>
</tr>
<tr>
<td>adjective (14)</td>
<td>79%</td>
<td>14%</td>
<td>7%</td>
</tr>
<tr>
<td>noun (23)</td>
<td>78%</td>
<td>18%</td>
<td>4%</td>
</tr>
</tbody>
</table>
Success with Verbs

Verbs
• Corpora are not POS-tagged
• May require complex searches: wild cards, multiple searches
• Potentially difficult

About a third involve alternatives
Q: use of verb attribute – attribute + to? Attribute something? [av Vb s]
O: to attribute something to something/someone. [s]

About a third are collocation queries
Q: what verbs can be used to lead effort? [ae Vb s]
O: make (+ adj) effort to; make every effort to; devote. [s]
Learning from Successful Students: Strategies for Novices

1. Formulate clear queries with items that can be searched.
2. Use both elicitation and verification queries.
3. Use alternatives to help narrow the query.
4. Use search words as probes to explore context/co-text.
References


