For citation:

Learning French from a Corpus: How and Why
Pan-SIG 2012
Hiroshima, Japan
Sunday, 2012 June 17, 11h15-11h45
Tom Cobb
Université du Québec à Montréal

HOW
• Almost everything on Lextutor will run French text (and TTS speech)
• Only small corpora are available, but they get big use from French teachers
  – The main 10 uses

WHY
• Using a corpus is part of a larger approach to language learning
  → Data Driven Learning (DDL)
• Is there a case for a broad deployment of DDL in French language training?
  Benefits? Obstacles?

IF TIME PERMITS

1. How do French teachers use Lextutor’s French corpus?

Tell me when you want to stop and go live

1. The obvious use – source of examples for the teacher
• Teacher finds examples to show students
  – Words
  – Structures
  – Discourse features
• Find sentences and texts for test questions

Display words, collocations, structures in classroom


Conclusion: most of “What it means to know a word” can be shown in a million-word corpus.
3. Give lexical info while reading

4. Encourage use of context before dictionary

5. Show if word is worth learning

6-7: Corpus as a while-writing learner resource

Expand the learner’s emerging text
• Via hook-up to a corpus of other writers’ texts
  — Again with payoff in strategy development

6. ConcordWriter

7. Data-Driven Error Analysis
8. Form-focus activities...
Ex, Mapping speech to text

9. Facilitate collaborative learning
• Since there are too many words to learn alone...

10. Facilitate quick-scope of a k-level
11. Give teacher a snapshot of a set of learner essays
   - Error patterns?
   - Recently learned words are coming through in production?
   - New structures coming through?

12. Get prep-phrases for a reading text

13. Scope out level of a reading text

Any research supporting all this?

- **COCONCORDANCE AS A READING RESOURCE**
  - Cobb, T., Graves, C., & Horst, M. *Can the rate of lexical acquisition from reading be increased? An experiment in reading French with a suite of on-line resources.* In F. Raymond & C. Cornaire (Eds.), *Regards sur la didactique des langues secondes.* Montréal: Éditions logique.

- **COCONCORDANCE AS WRITING FEEDBACK**

- **COCONCORDANCE INVESTIGATION OF LEARNER PRODUCTION**

- **COCONCORDANCE FOR SCOPING OUT A K-LEVEL**

II. Prospects for a French DDL

1. What is DDL?
2. Why does French need DDL?
3. What will facilitate the growth of DDL in French?
4. What obstacles will DDL face in French?
5. What DDL tools are French instructors using on Lextutor?
What is DDL?
(Data Driven Learning)
Or...
Didactique des langues!

DDL is...

- (Data Driven Learning)
- Where does DDL fit into the panorama of SLA (acquisition) models?
- Data-driven learning (DDL) is an approach to language learning within the input-driven paradigm, a paradigm that broadly emphasizes linguistic input and downplays intuition as the primary source of language acquisition. Given enough time and exposure, input theory proposes, learners gradually perceive the underlying lexical, grammatical and other patterns that are implicit in the language they hear and read around them. To this, the DDL approach contributes a nuance, which is that input in a second language (L2) will reveal its patterns more slowly and obscurely than is often required, but can do its work more quickly and accurately when reconfigured as data and run through certain kinds of language-analyzing computer programs. The most basic of these are variants of concordancing and frequency analyzers.

Why does French need DDL? 1

- Vast under-use of computing in LL generally relative to potential
  - and French in particular (as a major teaching language)
    - Partly related to lack of a coherent framework for CALL work

Why does French need DDL? 2

- Problems in the research with strict communicative LL
  - And related revival of Focus-on-Form and Language Awareness
    - (“within communicative context”)
  - Computational language analysis / activities / games
    - Are an effective and motivating way to work with form and raise awareness

Why does French need DDL? 3

- Growing awareness in the research of importance of automaticity in language processing
  - And discovery that few L2 learners achieve truly automated processes
  - Computers with timing ability are a major opportunity to develop automaticity
    - Also in-class methods like 3-2-1
Why does French need DDL?

Broader educational problem of preparation for life and work

• That computing is degenerating to context-intense interactions
  • (social and non-textual media)
  • Facebook etc
• ... and needs to be supplemented with more challenging and context-reduced interactions
  • Yesterday’s plenary themes

What will facilitate a French DDL?

Facilitations 1

• DDL is not language specific
  – A technology developed for one can work for others with minor adaptation
    • Just needs corpora
  – French or any Roman language can get a lot of start-up development for free from English

Facilitations 2

• French teachers have shown willingness to use a range of DDL activities
  – And participate in their development

Facilitations 3

• Lots of text available in French
  – Language Learners do not need the same highly designed corpora as corpus linguists do
  – We are only trying to learn the language
    • not describe / define / characterise it

What obstacles to French DDL lie waiting?
Obstacles 1

- Conservatism at the top of the enterprise
  - ‘Didactique des langues’ is a prisoner to (a) linguistics and (b) literature in France
- MA and PhD students must follow professors’ interests
  - Rarely involved with language teaching

Obstacles 2

- Unavailability of French corpora
- Expense of French corpora
  - Tend to be huge and expensive (Parole...)
  - Mainly written language, all adult NS (native speakers)
    - No e.g. graded reader corpus
    - Totally “unpedagogicalized”
  - Nothing like Paul Nation’s mighty familization of 14 k-lists out of the BNC
    - Despite good work on Lexique.org

Following Nation’s heroic cracking of the 100 million-word BNC into 14 k-family units...

Many things became possible on Lextutor
Especially with smallish corpora

Obstacles 2

- General lack (so far) of figures in French like
  - Paul Nation,
  - Rob Waring,
  - Batia Laufer,
  - Norbert Schmitt,
  - Mike McCarthy
- who are true applied computational linguists
  - Tireless fighters, bridging computational linguistics and pedagogy

Will DDL take root in French?

An interesting question
  - Teachers like it when they know what it is
  - Students like it
  - Learning results are confirmed in English
- Stay tuned...

Some papers to read in/about French


Some papers to read in/about French


- In P. Raymond & C. Cornaire, Regards sur la didactique des langues secondes. Montréal: Éditions logique. [Ici en français.]

- Applied linguistics is not linguistics applied (Parallel English-French text).

- All available at lextutor.ca/cv/
- This PPT at lextutor.ca/cv/ppts/jalt_2012_fr.ppt

-- COBB.TOM@UQAM.CA