

For citation:

Cobb, Tom (2012) Learning French from a Corpus: How and Why. In: Reinelt, R. (ed.) (2012) *Other Language Educators 2012 at JaltCALL and PanSIG* compiled for OLE by Rudolf Reinelt Research Laboratory, Ehime University, Matsuyama, Japan, p. 8 – 16.

1



L I T E R A C Y  
SIGNALS OF EMERGENCE



## 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



2

### 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

3

## I. How do French teachers *use* Lextutor's French corpus?

Tell me when you want to stop and go live

4

## 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

5

## Display words, collocations, structures in classroom

at [[Back \(keep settings\)](#)] [Colloc summary](#)  
FEMME with associated word homme on LEFT side: sorted 1 wds left of key [Dictionnaire: Fren\\_Eng](#) [Get](#)  
[50] [Keyword](#) equals [femme](#) [Assoc](#) homme [on left](#) [Sorted](#) 1 [wds left](#) [of key](#)  
[for more context](#)  
x. Les rapports ambigus d'un vieil homme et d'une FEMME en fleur. Une tragédie intimiste sur le  
culte du corps. Les canons de l'esthétique homme- FEMME tendent à se brouiller et glissent vers  
question de savoir s'il vaut mieux naître homme ou FEMME n'a jamais trouvé de réponse satisfaisan  
que le personnage principal, l'homme trompé par sa FEMME, employant maints stratagèmes pour faire  
l'opéra de Michaël Nyman, L'homme qui prenait sa FEMME pour un chapeau. OUVRAGE HYBRIDE Trouble  
fend, aux assises, un homme accusé d'avoir tué sa FEMME infidèle et l'ami de celle-ci avec un  
l'argent sur l'assistance. Face un homme ou une FEMME qui a 2 300 francs d'allocation, à la li  
secrète ? Le roman d'une pensée. Un homme aime une FEMME; il en est bouleversé; il s'essaie à com  
intersection de ces deux affaires, une homme et une FEMME : Alfred Sirven, ancien directeur des 'a  
si elle travaillait comme un homme, ou comme une FEMME : Elle croit ce qu'elle dit et dit ce qu'  
nts disposent des mêmes revenus : un homme et une FEMME mettant en commun un salaire identique v  
rait supérieur au SMIC que gagne un homme ou une FEMME qui va au travail chaque matin, a-t-il s  
freq=2) for femme (assoc-homme)



### 3. Give lexical info while reading

Concordance for equals SAS sorted by words left of key:

7 hits

001. affecté au bataillon de la brigade Special Air Service ( SAS) en août 1943, il est ensuite muté au régiment de ch

002. pour la circonstance. Cette société par action simplifiée ( SAS) sera constituée à 99 à par la Caisse française de dévelo

003. ouz l'avenant syndical (CA). Cette structure fait office de SAS pour "faciliter la sortie et assurer la solidarité et le

004. Rhône) : un arrêté autorisant la société Télépage France SAS à fournir le service téléphonique au public. Matignon :

005. s'agit ipse LE MONDE DES INITIATIVES l'indemnité d'insertion, un SAS vers l'emploi Expérience. Le réseau d'ores Interact est a

006. de formation aux métiers de l'impôt. "Parce qu'il manque un SAS entre les formations classiques et les besoins des entre

007. le Lyonnais de ses arrivants-ils à créer de véritables SAS autour de leurs activités de communication tout en conti

### 4. Encourage use of context before dictionary

Concordance for equals SAS sorted by words left of key:

7 hits

001. affecté au bataillon de la brigade Special Air Service ( SAS) en août 1943, il est ensuite muté au régiment de ch

002. pour la circonstance. Cette société par action simplifiée ( SAS) sera constituée à 99 à par la Caisse française de dévelo

003. ouz l'avenant syndical (CA). Cette structure fait office de SAS pour "faciliter la sortie et assurer la solidarité et le

004. Rhône) : un arrêté autorisant la société Télépage France SAS à fournir le service téléphonique au public. Matignon :

005. s'agit ipse LE MONDE DES INITIATIVES l'indemnité d'insertion, un SAS vers l'emploi Expérience. Le réseau d'ores Interact est a

006. de formation aux métiers de l'impôt. "Parce qu'il manque un SAS entre les formations classiques et les besoins des entre

007. le Lyonnais de ses arrivants-ils à créer de véritables SAS autour de leurs activités de communication tout en conti

### 5. Show if word is worth learning

Concordance for equals boule1 sorted by words left of key:

1 hit

001. Pendant plusieurs jours de suite, des lambeaux d'armée en déroute avaient traversé la ville. Ce n'était point de la troupe, mais des hordes débârdées. Les hommes avaient la barbe longue et sale, des uniformes en guenilles, et ils avançaient d'une allure molle, sans drapeau, sans régiment. Tous semblaient accablés, éreintés, incapables d'une pensée ou d'une résolution, marchant seulement par habitude, et tombant de fatigue s'ils qu'ils s'arrêtaient. On voyait surtout des mobillots, gens pacifiques, rentiers tranquilles, plant sous le poids du fusil, des petits mobillots alertes, faciles à l'épouvante et prompts à l'enthousiasme, prêts à l'attaque comme à la fuite; puis, au milieu d'eux, quelques culottes rouges, débris d'une division mourante dans une grande bataille; des

### 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

Concordance for equals REUSSIT sorted by words left of key:

13 hits

001. action, né en 1964, invente ce qui sera une des plus belles REUSSITES de marketing politique : la main de ferme et son e

002. sous-entend le succès que nous venons de composer. Cette REUSSITE, personne ne l'aurait osé possible, il y a seulement

003. oclairé toute sa vie par les coeurs. L'inspiration de cette REUSSITE, elle reçoit plus d'inspirations, haïrises sous une

004. e cap des 500 000 ventes par mois est français. Devant cette REUSSITE, les chaises, à leur tour, ont lancé des litanies

005. il est "un des rares à connaître les complex exacts. Cette REUSSITE lui permet de devenir directeur général en 1994. Ma

006. REUSSIT, ses traits sans différencie expliquent aussi cette REUSSITE. Le Mondial 1998 et l'enthousiasme qu'il a créé à B

007. ces dernières années, laissent perplexe sur ses chances de REUSSITE. PAR DANIEL VERDET LES MÉNAGES D'UN SOLDAT d'Alain

008. 84. Deuxième élément justifiant cette évolution, le taux de REUSSITE au bac 3 diminue. Alors qu'il atteignait des sommets

### 7. Data-Driven Error Analysis

Concordance for equals REUSSIT sorted by words left of key:

13 hits

001. action, né en 1964, invente ce qui sera une des plus belles REUSSITES de marketing politique : la main de ferme et son e

002. sous-entend le succès que nous venons de composer. Cette REUSSITE, personne ne l'aurait osé possible, il y a seulement

003. oclairé toute sa vie par les coeurs. L'inspiration de cette REUSSITE, elle reçoit plus d'inspirations, haïrises sous une

004. e cap des 500 000 ventes par mois est français. Devant cette REUSSITE, les chaises, à leur tour, ont lancé des litanies

005. il est "un des rares à connaître les complex exacts. Cette REUSSITE lui permet de devenir directeur général en 1994. Ma

006. REUSSIT, ses traits sans différencie expliquent aussi cette REUSSITE. Le Mondial 1998 et l'enthousiasme qu'il a créé à B

007. ces dernières années, laissent perplexe sur ses chances de REUSSITE. PAR DANIEL VERDET LES MÉNAGES D'UN SOLDAT d'Alain

008. 84. Deuxième élément justifiant cette évolution, le taux de REUSSITE au bac 3 diminue. Alors qu'il atteignait des sommets





## What is DDL ?

(Data Driven Learning)

Or...

Didactique des langues!

31

## 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.

<<<UNPACK in 2-3 SLIDES

32

## Why does French need DDL?

33

## 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

34

## 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

35

## 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

36

## Why does French need DDL? 4

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

37

## What will facilitate a French DDL?

38

## 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

39

## Facilitations 2

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

40

## 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

41

## What obstacles to French DDL lie waiting?

42

## 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

43

## 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 familiarization of 14 k-lists out of the BNC
      - Despite good work on Lexique.org

44

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

45

## 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

46

## 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...

47

## Some papers to read in/about French

- Ovtcharov, V. , Cobb, T. & Halter, R. [La richesse lexicale des productions orales : mesure fiable du niveau de compétence langagière.](#) *Revue Canadienne des Langues Vivantes*, 63(1), 107-125.
- Cobb, T. (2009). [Review of Se donner le mot: Trousse pédagogique.](#) *Revue Canadienne des Langues Vivantes*, 65(4), 645-647.

48



## Some papers to read in/about French

- Cobb, T. (2007). *My Word Coach* (Jeu vidéo). Montréal: Ubisoft Inc. [[Journal l'UQAM, Jan 08](#)]
- Cobb, T., Greaves, C., & Horst, M. (2001). **Peut-on augmenter le rythme d'acquisition lexicale par la lecture ? Une expérience de lecture en français appuyée sur une série de ressources en ligne.**
- In P. Raymond & C. Cornaire, *Regards sur la didactique des langues secondes*. Montréal: Éditions logique. [Ici [en français.](#)]

49

- [Applied linguistics is not linguistics applied](#) (Parallel English-French text).
- \*\*\*\*\*
- All available at [textutor.ca/cv/](http://textutor.ca/cv/)
- This PPT at [textutor.ca/cv/ppts/jalt\\_2012\\_fr.ppt](http://textutor.ca/cv/ppts/jalt_2012_fr.ppt)

– COBB.TOM@UQAM.CA

50