Introduction générale

SÉMANTIQUE HISTORIQUE

Procédures formelles pour l'analyse de corpus de textes anciens

A. Héritage et situation

A1. Le historiens et la sémantique

A2. Langues, sociétés, significations

A3. Conjoncture actuelle

A4. Exemples

B. Problèmes et possibilités

B1. Les caractères spécifiques des corpus de textes anciens et la sémantique historique

B2. Renverser la perspective ?

B3. Nécessité et difficultés de la formalisation

B4. Exemples

C. L'école d'été de Ciutadella

C1. Définition de l'objectif

C2. Plan de l'école d'été

A1 Les historiens et la sémantique

- Ernst Bernheim, *Lehrbuch der historischen Methode*, Leipzig, 1889. « Besonders ist es der Wandel in der Bedeutung und Anwendungsweise der Worte, worauf wir zu achten haben. » (éd. 1908, p. 578).
- Charles Seignobos, *Introduction aux études historiques*, Paris, 1897. « Pour interpréter Grégoire de Tours, ce n'est pas assez de savoir en général le latin ; il faut encore une interprétation historique spéciale pour adapter cette connaissance générale au latin de Grégoire de Tours. » (5e éd. p. 122).
- Marc Bloch, *Apologie pour l'histoire*, 1943. « Certains de nos aînés, comme Fustel de Coulanges, nous ont donné d'admirables modèles de cette étude des sens, de cette « sémantique historique ». Depuis leur temps, les progrès de la linguistique ont encore aiguisé l'outil. » (éd. 1993, p. 174).
- Jacques Le Goff, « Introduction » à l'édition précédente, « L'historien s'attachera… à l'étude des sens, à la « sémantique historique », dont il faut souhaiter la renaissance aujourd'hui ». (ibid. p. 29).
- Création de la revue *Historische Semantik*, Göttingen, 2003.
- Résultat : aucune synthèse, aucun manuel ; le recours aux dictionnaires est toujours l'ultima ratio, utilisation massive et irréfléchie de prétendues « traductions ».

A2 Langues, sociétés, significations

- Premières réflexions sur langue et société : Friedrich August WOLF (1759-1824), *Prolegomena ad homerum* (1795), et surtout Wilhelm von HUMBOLDT (1767-1835) : *Über die Kawi-Sprache auf der Insel Java*, 1836. [trad. fr. *Sur la diversité de construction des langues et leur influence sur le développement de la pensée humaine]*
- Disparition de ce type de recherche au profit de la philologie, qui s'attache à l'évolution des formes, en particulier à la phonétique. L'histoire se détourne peu à peu d'une visée rationnelle, pour sombrer dans les mythes nationalistes.
- Apparition de la « sémantique » à la fin du 19e siècle (Michel BRÉAL 1832-1915) : « histoire des mots » ; chaque mot est considéré isolément, un mot = une chose ! Perspective purement atomiste.
- Découverte de la paire Wortfeld - Sinnbezirk : Jost TRIER (1894-1970)
  a) nécessité de considérer **trois** ensembles : la société, les concepts (zones sémantiques) et les mots (champs lexicaux)
  b) nécessité de considérer les **liens** et les **réseaux**, pas les éléments
  c) nécessité d'une observation dans des périodes successives, privilège de l'**analyse des changements.**
- La linguistique contemporaine s'est enfermée dans l'idée d'une structure indépendante et intemporelle. La sémantique actuelle peine à retrouver les découvertes de Trier, et ne s'en rapproche que lentement. La plupart des outils de « traitement automatique des langues » partent de l'idée que le sens des mots est connu !! (« wordnet », « ontologies », « web sémantique », etc).

A3 Conjoncture actuelle

- ***Indifférence absolue*** des historiens, affaiblissement continu de la connaissance de base des langues (anciennes et contemporaines), ignorance totale des premiers éléments de philologie.
- Autoenfermement de la linguistique théorique dans le ***phantasme*** d'un objet clos sur lui-même. Langue = objet pur, en dehors de toute société et de tout temps.
- ***Survie laborieuse*** de la lexicographie historique traditionnelle.
- Explosion des méthodes pratiques de recherche et de manipulation des « textes » sur internet, énormes enjeux financiers ; anglo-américain comme seul objet, aucun souci du sens, supposé par principe connu (wordnet) ; encore moins de souci (!) d'une quelconque évolution des sens. Tout cela dans l'« océan digital ».
- Quelques « trouées » informatiques surprenantes, ponctuelles (ngram-viewer ; htt://corpora.informatik.uni-leipzig.de) ; de plus en plus de corpus numérisés accessibles.

A4 exemples

- Pour fixer les idées, quelques exemples tirés du français contemporain. Ou : comment la définition d'un objet ne rend pas compte du sens du mot.
- Deux mots pour le même objet : *cotisations* ou *charges* ; *contributions* ou impôt ; *héritage culturel* ou patrimoine.
- Un même mot, avec des sens opposés selon le contexte et les locuteurs : *individualisme*, *dieu*, *sécurité* (sociale ou « des biens et des personnes »)

B1 Les caractères spécifiques des corpus de textes anciens et de la sémantique historique

- CORPUS
  *** corpus clos, longueur fixe et restreinte
  *** textes non-contemporains les uns des autres (hétérogénéité intrinsèque)
  *** tous les textes dans le domaine public
- SÉMANTIQUE
  situation diamétralement opposée à la situation des textes actuels :
  ***sociétés contemporaines, même aire culturelle : bilinguisme, vérifications ; parole = contact direct. difficultés subsistantes !
  ***sociétés anciennes : écrit exclusivement, références concrètes disparues, système de relations inconnu… > **textes opaques**
- J.F. Champollion, 1836
  comment a-t-il fait ?
  ➔ véritable défi, front pionnier !

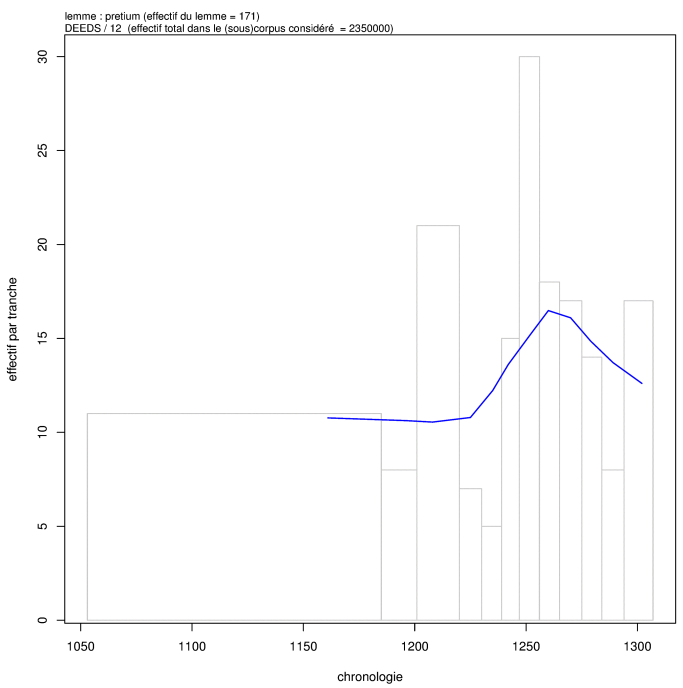
B2 Renverser la perspective ?

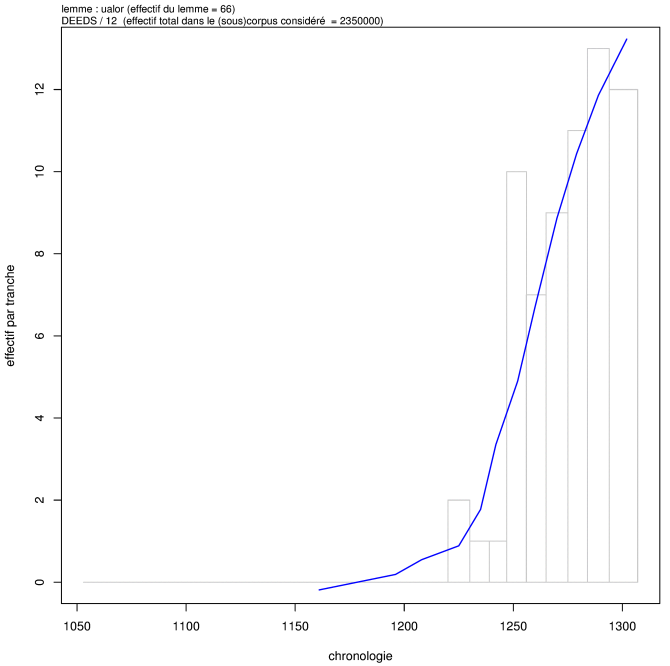
- Champollion n'était pas préoccupé de définir les termes, mais de comprendre la langue, i.e. le codage graphique.
- Nous devons tenter d'identifier un codage sémantique qui nous échappe en très grande partie.
- Il vaudrait peut-être mieux **se préoccuper de comprendre les textes anciens avant de tenter d'y « chercher des informations »**. Cela suppose une critique radicale de la notion sacro-sainte de « problématique »

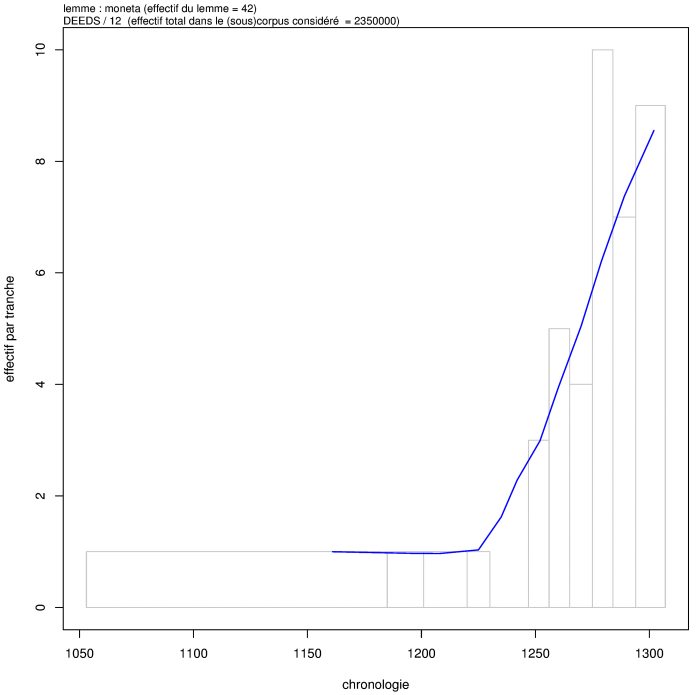
B3 Nécessité et difficultés de la formalisation

- Mots rares : quelques occurrences, lecture directe.
- Mots fréquents : toute tentative d'embrasser des centaines ou des milliers d'occurrences est entachée de subjectivité, résultats insuffisants, incontrôlables, souvent faux > les mots les plus fréquents sont les plus difficiles.
- Or toute manipulation informatique suppose des caractères nettement identifiables : catégories, types > nécessité de définitions explicites. Les objets sont donc munis de marqueurs formels homogènes > les structures mises au jour résultent de **présupposés explicites**. Corpus + méthodes formelles > résultats contrôlables, discutables.
- Les résultats dépendent donc très largement de la structure du ou des corpus et des choix de catégorisation a priori : comment choisir ???

B4 Exemple 1 ***ualor***

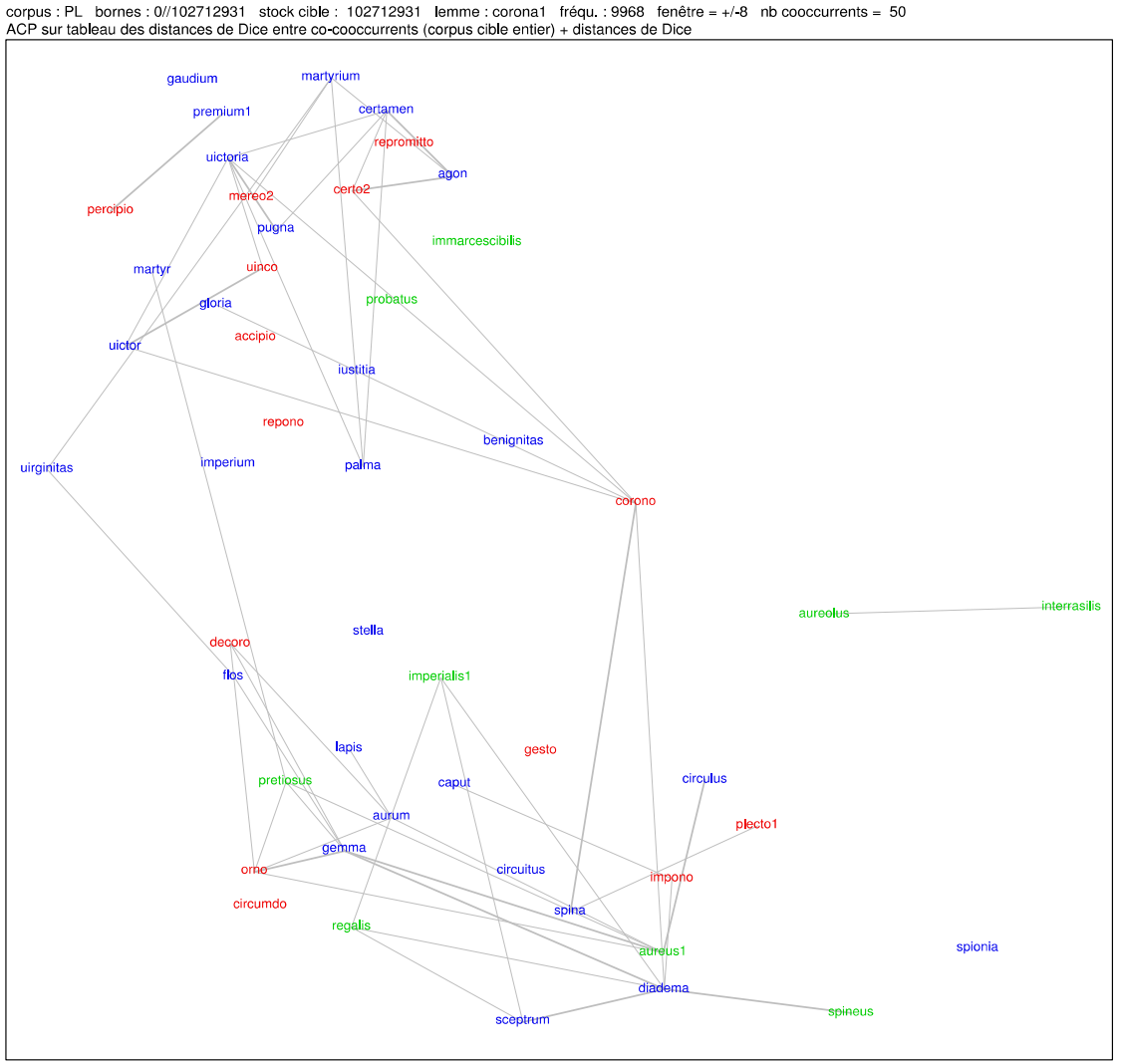






- Corpus DEEDS : 10000 chartes anglaises datées, rassemblées par Michael Gervers.
- ***ualor*** n'est pas attesté avant 1220 mais connaît une croissance rapide à partir de 1250
- ***pretium***, lui, est présent dans tout le corpus, sans tendance repérable (*denarius* : idem)
- ***moneta*** fait une timide apparition au 12e, mais ne se développe nettement que dans la seconde moitié du 13e.
- Que signifie alors l'apparition de ce terme (création médiévale) ?

B3 Exemple 2 ***corona***



- Corona dans la PL, près de 10000 occurrences.
- Répartition des cooccurrents (lexicogramme).
- Au moins trois ensembles distincts apparaissent : avant tout, la *corona martyrum*, en face, la *corona christi* (couronne d'épines), dans l'entre-deux la corona imperialis.
- S'agit-il d'une structure propre à la PL ou d'un élément du système de représentation médiéval ?

C1 Définition de l'objectif

1. créer des corpus de textes historiques (ici : principalement latins), convenablement organisés ;
2. prendre en mains des outils (procédures informatiques) apportant une aide à la perception du sens des mots et des textes de ces corpus.

ATTENTION :
*** une aide qui vient **en plus** des méthodes traditionnelles, sans les remplacer.
*** ces opérations n'ont **rien d'automatique**, il faut les comprendre de l'intérieur, être capable de les modifier, et au besoin d'en inventer de nouveaux. Nécessité de logiciels libres, « *mettre les mains dans le cambouis* ».
*** aucune interprétation ne peut être interne aux textes, pas de sémantique sans une connaissance approfondie de ce qu'on sait des structure sociales.

C2 Plan de l'École d'été

1. le système linux et les outils nécessaires à la construction de corpus
2. étapes de construction d'un corpus
3. outils de base pour la manipulation d'un corpus de textes
4. statistique lexicale et procédures numériques
5. applications aux dictionnaires et à la lexicographie

Importance des exercices : 3 demi-journées complètes + 3 partielles (= au moins 1/3)

QUESTIONS ?