Semantic characteristics and use of some specific cross-disciplinary lexemes in French scientific papers

Cristelle Cavalla

Francis Grossmann

LIDILEM, EA 609 – Grenoble III

Comparing forms academic organisation, cultural identities and scientific discourse across nations and disciplines

Paris, November 15-16 2004

Introduction (1)

- Work is carried out within the scope of the "Emergence" project: phraseology of the scientific paper, in academic cross-disciplinary lexicon (common to various disciplines).

- Involves a large and heterogeneous lexical field concerning methods, theoretical approach, problematics, evaluation of results, etc.

Introduction (2)

- Some co-occurrences V+N, represent a kind of scientific emblem, e.g. : *faire une hypothèse* (*to hypothesize* or *to theorize that*)

  - ➤ But these collocations present polysemic significations;
  - ➤ And they are used differently according to epistemological traditions, national cultures, scientific forms of writing…

- For this presentation, we will limit ourselves to investigating the use of a few terms in French : *hypothèse, thèse, théorie* and *postulat*.

3

Our objectives

Observe verb-noun co-occurrences (**V + N**) for the four selected nouns (*hypothèse, thèse, théorie, postulat*) in order to pinpoint their use in scientific papers:

- see if they share semantic characteristics so as to explain lexical associations with these nouns

- see if they vary according to specific semantic features or dimensions

- observe the various use of the same lexeme

Method (1)

- **Corpus study**: Rinck (Linguistics) + Per Hermès (Human Sciences) + KIAP (various disciplines) = about 1700 000 words in all.

- **Search and classification of noun+verb co-occurrences** through concordances

- **Lexicographical support (TLF) in order to:**
  - disambiguate meanings and select useful meanings in relation to the scientific discourse;
  - have an auxiliary corpus to check collocations

Method (2)

■ Selection of fairly **broad dimensions** in order to partially neutralise the freezing process inherent to collocations: e.g. *émettre une hypothèse/énoncer une hypothèse*

→ no significant distinction in use.

Study of 'hypothèse'(1)

TLF: in philosophy and sciences, the specialised meaning divides into two specific acceptations:

a) **proposition that is accepted** regardless of its truth and on the basis of which a given set of propositions is derived. Syn. *Principle*

b) **proposition (or set of propositions) that is temporarily accepted** in order to explain facts, natural phenomena and has to be checked through deduction or experience. Syn. *Conjecture*.

Note: used in ordinary language as a synonym of *supposition*

Study of 'hypothèse'(2)

4 main dimensions in the corpus

A. Formulation:

– Typical verbs: *émettre (16), formuler (9), poser (9), présenter (4), introduire (2) avancer (17), énoncer (1), proposer (6)*

B. Elaboration:

– Typical verbs: *faire (113), effectuer (4), former (1)*

Study of 'hypothèse'(3)

C. Verification

– Typical verbs: *vérifier (6), tester (35), confirmer (12), corroborer (3), mettre à l'épreuve (3), valider (6), examiner (3), infirmer (4)*

D. Argumentation

– Typical verbs: *soutenir (2), étayer (3), conforter (2), défendre (9), appuyer (1), renforcer (4), justifier (4): discuter (3), légitimer (1) opposer (1), privilégier (2), récuser (1), soutenir(1)*

Study of 'hypothèse'(4)

■ **Note**: frequent merging of two different meanings of the verb *faire*: *« Le premier programme scientifique en linguistique qui a fait l'hypothèse du langage naturel comme système formel »*

- 1. *Faire* is a support verb (see $oper_1$ in MTT): grammatical tool used to make up for the missing verb "*Hypothéser* "

- 2. *Faire* as a verb of elaboration: *faire* = build, make, produce.

- → because of this ambiguity, this collocation has not been taken into account in the dimensions (and is not included in the following graph).

Main dimensions for 'hypothèse'
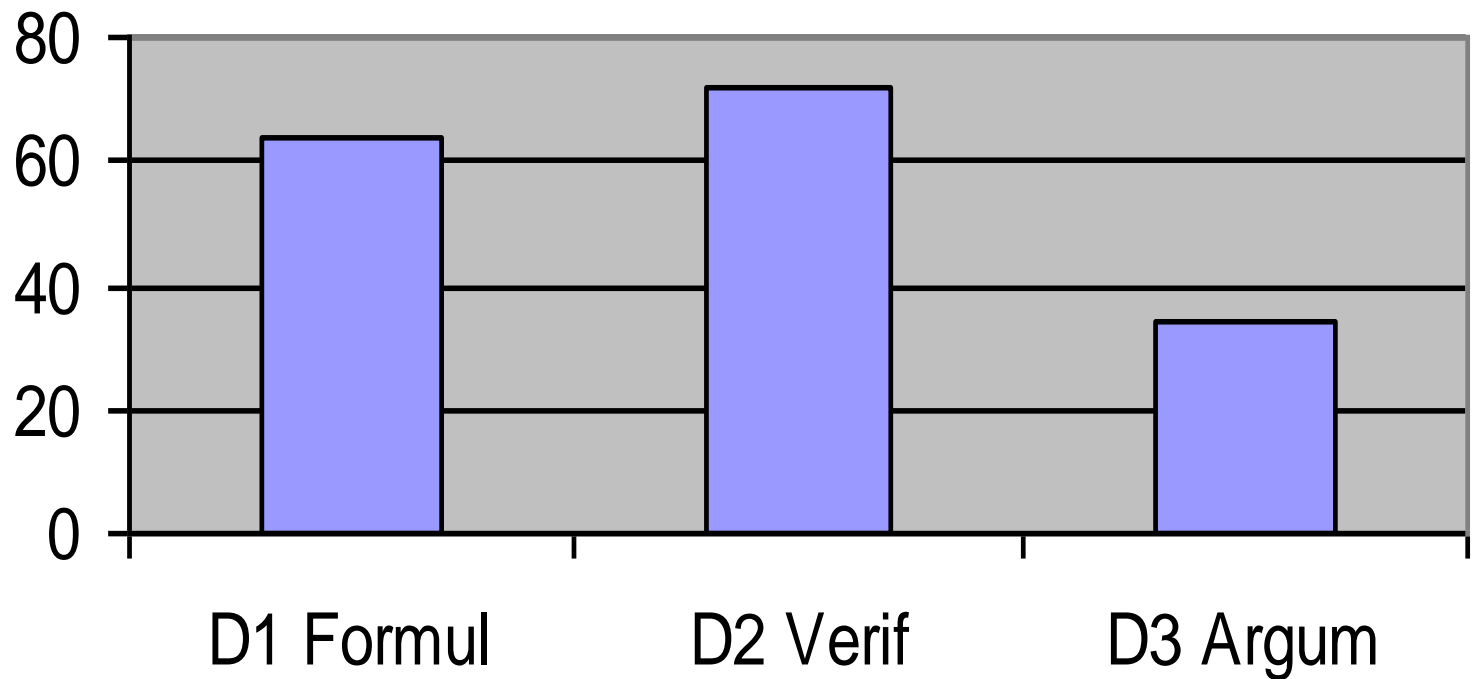
Study of 'postulat' (1)

TLF: in geometry and sciences, the specialised meaning divides into three specific acceptations:

a) Proposition that has to be taken as a premise of a demonstration even if it is not obvious and has not been proven.

Study of 'postulat' (2)

b) Proposition that is part of the axiomatics taken as a premise of a hypothetico-deductive system. Syn. *Axiom*

c) Principle that has not been proven but is accepted as a basis for research or theory.

Note: also used in ordinary language for an implicitly accepted representation on which a thinking system is based.

Study of 'postulat' (3)

Very few occurrences, only 13 in the corpus (because of the evident competition with the verb *postuler*); *postulat* seems to be seldom used, particularly in human sciences.

Affiner un postulat (1), Avancer un postulat (1) Ajouter un postulat (1), Appliquer un postulat (1), **Confirmer un postulat** *(1), Décliner un postulat (1), Implémenter un postulat (1,* **Infirmer un postulat** *(1), Réactiver un postulat (1), Redoubler un postulat (1), Ruiner un postulat (1),* **Valider un postulat** *(1), relativiser un postulat (1)*

→**dimensions are difficult to define**

→ **In HS: very close to *hypothèse (same semantic features as /premise/ or /starting point/)***

Study of 'postulat' (4)

- **Investigation of a complementary corpus: Frantext (category "essays" from 1900 to 2000):**

  - *Admettre le postulat* : 20 occurrences; *rejeter le postulat*: 14; *nier* : 1;

  - *Confirmer* : 2, *infirmer* : 2; *démontrer* : 6;

  - *Poser (comme) postulat* : 2.

→ **Fairly strong argumentative dimension in view of this new corpus; in HS sometimes used with a neg. meaning:** *Mais ce n'est là que <*postulat*> et oubli des vérités*   **(Foucault)**

Study of 'thèse' (1)

TLF: *Proposition or theory that is considered to be true and supported with arguments in order to prevent any objections.*

- *Note 1: the meaning of "thèse" as doctorial dissertation has not been taken into account.*

- *Note 2: this meaning is not specialised even if used in sciences.*

Study of 'thèse' (2)

- **Two main dimensions (with the same reservations due to the small number of occurrences):**

    A. Argumentation: *Appuyer (3), contester (1), défendre (6), réfuter (1), soutenir (1), rejeter (1), exclure (1), privilégier (1), exclure (1)*

    B. Formulation: *Avancer (1), exposer (1), formuler (1), énoncer (1)*

- **The argumentative dimension ranks first for *thèse* (as compared with formulation; this is the opposite of what has been observed for *hypothèse*); it seems logical, in view of its lexicographical definition.**

Study of 'théorie' (1)

Specialised meaning in sciences:

 A. Intellectual, hypothetical and synthetic construction, organised as a system and verified according to an experimental protocol;

 B. Set of laws forming a coherent system used as a basis for a science or to account for facts.

Note: unlike *hypothèse* and *postulat* that have different meanings according to the scientific field in which they are used, there is no evidence of polysemy for *théorie* in sciences.

Study of 'théorie' (2)

Corpus:

A. Elaboration: *construire (3), esquisser (3), fonder (2), établir (1), édifier (2), élaborer (2), établir (1), jeter les bases (2), déboucher sur (1), produire (1)*

B. Formulation: *exposer (3), introduire (1), présenter (2), proposer (5), décrire (1)*

C. Verification: *confirmer (1), évaluer (2), falsifier (1), tester (1), vérifier (1)*

D. Argumentation: *accepter (1), admettre (1), adopter (1), défendre (1), démonter (1), justifier (1), mettre en cause (1),reconsidérer (1), rejeter (1), tourner en dérision (1).*
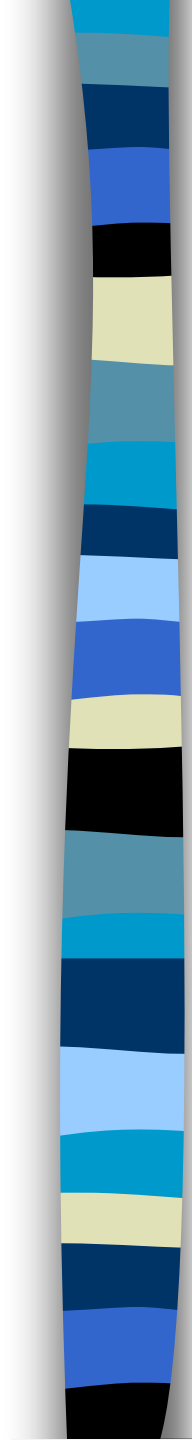
Main dimensions for 'théorie'
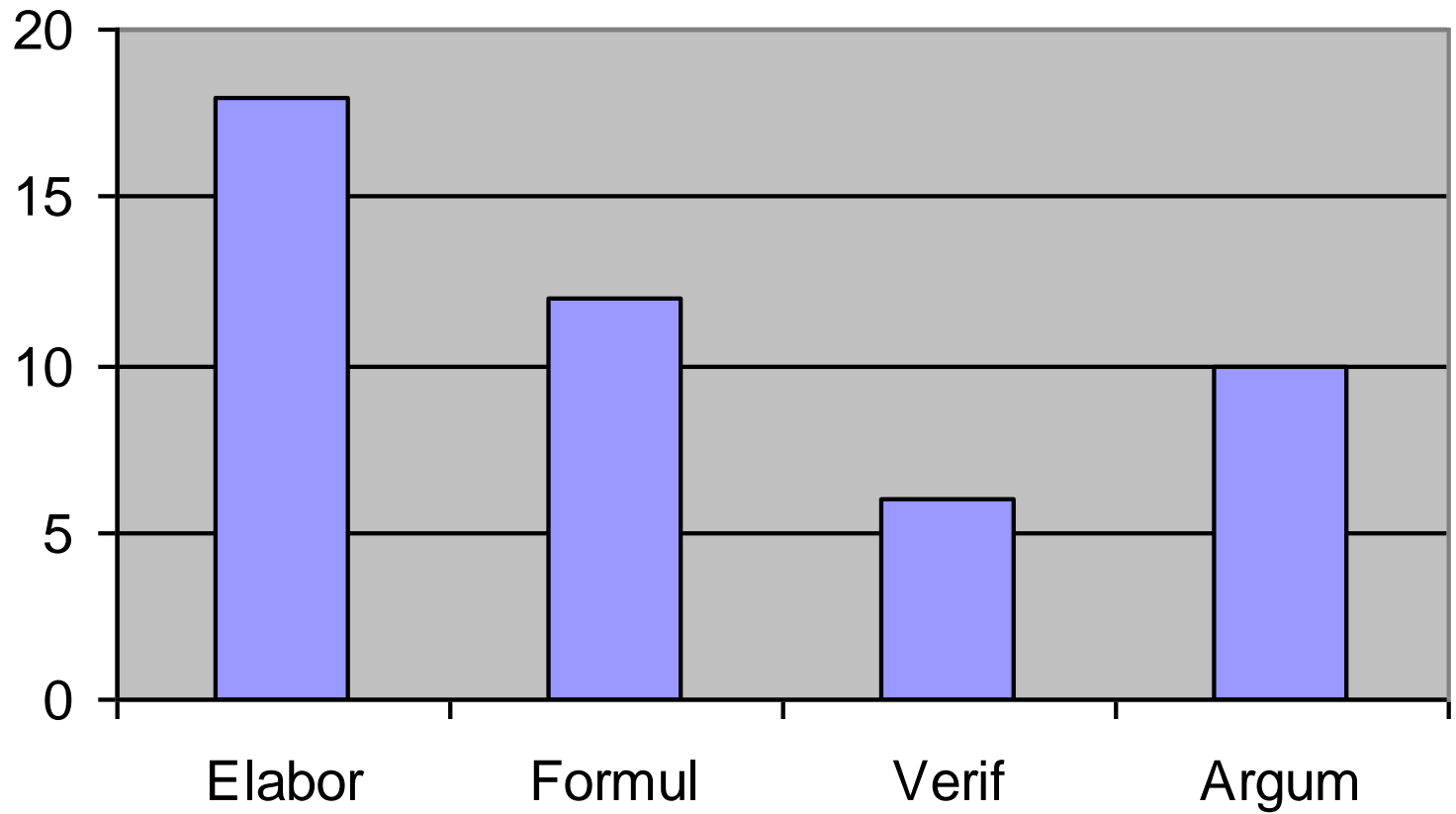
Summary

- Few verbs are found in co-occurrence with the 4 nouns, but usage shows preferences rather than exclusions:
  - *émettre* or *énoncer une hypothèse* rather than *exposer une hypothèse*
  - *exposer une théorie* rather than *énoncer une théorie*

- There is a common "semic core" but the selection of verbal co-occurrents sometimes differs according to the nouns (ex. **Formulation,** see example above)

- The diversity of use is partially linked to the polysemy of the terms in the scientific field (or non-scientific field, e.g. *thèse*)

	Elaboration	Verification	Formulation	Argumentation
hypothèse	low frequency (corpus: no occ. except with 'faire' TLF: *échafauder, construire*)	high frequency	–	low frequency (corpus: *soutenir, discuter, récuser…*)
postulat	–	low frequency (?)	–	average frequency
thèse	–	low frequ. (corpus: no occ. TLF: confirmer, infirmer)	average frequency	**high frequency**
théorie	**high frequency**	low frequency	average frequency	average frequency

Semantic profiles of the scientific nouns under study according to the dimensions found in the corpus and in the TLF

Semantic explanation (1)

The differences that have been found can be clarified thanks to a semantic analysis of the four terms:

- *Hypothèse* and *postulat*:
  - /assertion/ /premise/ or /starting point/; see collocations such as *partir de l'hypothèse/du postulat*

- *Théorie* and *postulat*:
  - /system/ or /organisation/

- *Théorie:*
  - /elaboration/

Semantic explanation (2)

– *Hypothèse*:

- /true or false/

– *Postulat:*

- /accepted/ /true by definition/; collocations such as *confirmer le postulat* not frequent but possible

– *Thèse:*

- /assertion/ /to be advocated/ /true for its advocates/

Conclusion (1)

■ Thanks to this multidimensional approach it is possible to anticipate the use of verbs that can be associated with different classes of academic nouns:

  Hypothèse and *Postulat*: the "FORMULATION" dimension makes it possible to use verbs such as "*formuler, énoncer, invoquer, expliciter, expliquer, rendre compte de, définir;* however, it should be possible to explain restrictions such as "**Affirmer l'hypothèse/le postulat, *dire l'hypothèse/le postulat*

■ This approach also explains discrepancies in frequency: the "Elaboration" dimension is preferred for *théorie* but it is possible for *hypothèse* (even if it has not been found in the corpus: *construire une hypothèse*)

Conclusion (2)

- However, a more refined semantic approach helps to understand some specific points:

  e.g. the relation to the truth value can explain the discrepancies in frequency of verbs such as *vérifier, confirmer* with *hypothèse* (which can be true or false) and "*postulat*" (which is true by definition); see also: *avancer une hypothèse* rather than *avancer un postulat.*

- Solutions: more dimensions ? This might reduce the output by multiplying micro-classes. Dimensional approach + secondary semantic filter ?

Conclusion (3)

Differences in use:

- some of the lexemes can be used whatever the field
  - *thèse* and *théorie*

- others can only be used in some fields
  - except in NLP papers, *postulat* is seldom used in HS, and when it occurs its meaning is close to that of *hypothèse* or it is used in a negative sense (an assumption that is implicitly accepted but should be rejected).

Conclusion (4)

Perspectives

– It would be interesting to take into account enunciative aspects, e.g. modalisation with adjectival collocations :

  - *L'hypothèse est juste, insuffisante, séduisante, originale, saugrenue, réfutable, réaliste, recevable, économique…*

– It would also be useful to compare fields more accurately:

  - e.g. use of *hypothèse* in medicine *vs* linguistics; it seems that there is a significant difference in the use of *tester*, *vérifier* (more experimental approach in medicine than in linguistics).

– Multilingual approach: compare with other corpus studies in other languages to highlight differences

  - e.g. when the French say "*Émettre l'hypothèse que*…" the English might say "*to hypothesize*" or "*to theorize that*…".

Takk for oppmerksomhet !

Merci de votre attention !
Thank you for your attention !