Do biology and linguistics belong to the same academic discourse?

* A corpus-based contrastive analysis

Céline Poudat (celine.poudat@unice.fr) and Peter Follette (follette@unice.fr)

Université de Nice - Sophia Antipolis

GERAS Conference, March 18-20, 2010
Outline

- Biology and linguistics: two different academic worlds and practices
- Corpora
- Article structure
- Contrasting the two corpora
Corpora

Biology

Linguistics
• 90 research articles (2004-2009) in Molecular Biology
• 45 natives / 45 non-natives (matched - journal/theme)
• 36 different journals (9 unpublished)
• Removal of figures, tables, references
Linguistics

- A main corpus of 103 articles: **LING103**
- A working corpus (subset): **LING51**
  - 51 articles, 62 authors
  - 6 journals
  - 28 natives / 23 non-natives
  - 9 articles with multiple authors
- Removed:
  - Abstracts
  - Examples, figures, tables
  - References
Article structure

IMRAD vs IMRAC
IMRAD Structure in Biology

• Introduction (What was known about the topic before the study? What do we need to know?)
IMRAD Structure in Biology

• Introduction (What was known about the topic before the study? What do we need to know?)

• Materials & Methods (How were the experiments performed?)
IMRAD Structure in Biology

• Introduction (What was known about the topic before the study? What do we need to know?)

• Materials & Methods (How were the experiments performed?)

• Results (What did we do and what did we see?)
IMRAD Structure in Biology

• Introduction (What was known about the topic before the study? What do we need to know?)

• Materials & Methods (How were the experiments performed?)

• Results (What did we do and what did we see?)

• Discussion (What does it mean? What do we know now?)
Types of articles in Biology

- Hypothesis-driven (Testing the role of gene X in process Y)
- Descriptive (Discovery of a new structure)
- Data-driven (Microarray analysis)
- Method-driven (Development of a new method of labeling proteins)
Introduction
- Introduce the topic
- Review items of previous research
- Indicate a gap, Continue a tradition, Counter-claim
- Introduce the current research study

Material
- Corpus: corpus-based or corpus-driven
- Created examples
- Review articles: main references in the field
- Empirical articles: specific measures or methods used in data analysis

Analysis (RAD)
- The three ‘moves’ scarcely match articles sections
- Typological presentations of the results are more common

Conclusion
- Answer the research question
- State importance and limitations of the findings
- Announce further research
**Experimental articles (IMRAD-like)**
Ex: Schmid: 'Presupposition can be a bluff': How abstract nouns can be used as presupposition triggers

- Introduction
  - Background
  - Indicating a gap

- Material
  - Corpus data
  - Reliance and attraction measures

- Analysis (RAD)

- Conclusion
  “Apparently, presupposition can indeed be a bluff.”

**Claim articles (Essays)**
Ex.: Lindblom: Cooperating with Grice. A cross-disciplinary metaperspective on uses of Grice's cooperative principle

- Introduction
  - Background
  - Counter-claim

- Material
  - Review of Grice

- Proposal
  - Metaperspective

- Analysis (RAD)

- Conclusion
  • Summary, Answer
Contrastive exploration
Research directions

Main contrasts

- Biology vs. linguistics
- Native vs. non-native

Observations and Methods

- Domain-specific themes
- Repeated segments
- Style and morphosyntax
- Scientific vocabulary: epistemic/ signalling nouns
Reprinted segments, or prefabs

<table>
<thead>
<tr>
<th>Biology</th>
<th>Linguistics (LING103)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Previous research</strong></td>
<td><strong>Previous research</strong></td>
</tr>
<tr>
<td>- it has been shown that (16 occ.)</td>
<td>- the propositional content of the (23 occ.)</td>
</tr>
<tr>
<td>- previously been shown to (10)</td>
<td>- differences in the use of (10)</td>
</tr>
<tr>
<td>- has been proposed that (10)</td>
<td>- as in the case of (10)</td>
</tr>
<tr>
<td><strong>Methods</strong></td>
<td><strong>Methods and objects (more variation)</strong></td>
</tr>
<tr>
<td>- according to the manufacturer’s ‘(instructions, protocol) (19)</td>
<td>- the propositional content of the (23 occ.)</td>
</tr>
<tr>
<td>- was used as a loading control (10)</td>
<td>- differences in the use of (10)</td>
</tr>
<tr>
<td>- experiments were carried out (10)</td>
<td>- as in the case of (10)</td>
</tr>
<tr>
<td><strong>Findings</strong></td>
<td><strong>Findings</strong></td>
</tr>
<tr>
<td>- We conclude/found that the (28)</td>
<td>- we found (11)</td>
</tr>
<tr>
<td>- we have shown that (14)</td>
<td>- have found (≠)</td>
</tr>
<tr>
<td>- These data suggest that (20)</td>
<td>-</td>
</tr>
<tr>
<td>- These results suggest/indicate that (33)</td>
<td>-</td>
</tr>
</tbody>
</table>
# Repeated segments, or prefabs

<table>
<thead>
<tr>
<th>Biology</th>
<th>Linguistics (LING103)</th>
</tr>
</thead>
<tbody>
<tr>
<td>• Focus</td>
<td>• Focus</td>
</tr>
<tr>
<td>• Textual</td>
<td>- is important to note that (11)</td>
</tr>
<tr>
<td>- as described in Materials and Methods (19)</td>
<td>- It is interesting to (10)</td>
</tr>
<tr>
<td>- As shown in Figure (32)</td>
<td>- It should be noted that (10)</td>
</tr>
<tr>
<td>- and data not shown (28)</td>
<td>- I-Focus: I would like (17)</td>
</tr>
<tr>
<td>• Discourse organizers</td>
<td>• Textual</td>
</tr>
<tr>
<td>- on the other hand (36)</td>
<td>- of this paper is to (12)</td>
</tr>
<tr>
<td>- At the same time (15)</td>
<td>- As we have seen (11)</td>
</tr>
<tr>
<td></td>
<td>- as can be seen (20)</td>
</tr>
<tr>
<td></td>
<td>• Discourse organizers</td>
</tr>
<tr>
<td></td>
<td>- on the other hand (157)</td>
</tr>
<tr>
<td></td>
<td>- on the one hand (56)</td>
</tr>
<tr>
<td></td>
<td>- at the same time (42)</td>
</tr>
</tbody>
</table>
Specificities: style and morphosyntax
+ linguistics, - biology

- **Determiners**: the, a, an,
- **Be-forms**: is, are, be, to be, it is (vs. was)
- **Pronouns**: I, YOU, it, SHE (her), he (his), they (their)
- that, this, who, which, what, how
- there
- Not, is not
- **Modals**: can, SHALL, may, must, would, should
- **Aux. do, Future will**
- to, as, more, such, own
- About
- But, given connective (184 vs. 15)

Common, Rare or Absent (in biology)
More common in biology
Specificities: scientific terms

- linguistics, - biology

- biology, - linguistics

- focus, interpretation, account, theory, examples, view, use of, relation, rules, question, claim, know, constraints, knowledge, research, in terms of, scope, discussion, section, perspective, distinction, relevant, paper, say, cf, way, particular, case, fact, take, structure, point, reanalysis, scientific

- showed, assay, was used, complex, observed in, measured, compared, required for, results, rate, tested, the absence, induction, effects, suggesting that, reaction, indicating
Linguistics vs. Biology
The pronoun “we”
The pronoun “we”
The pronoun “we”
En cas de recouvrement, dragged légèrement un mot pour le rendre visible. Maintenir la souris sur un point quelconque d'un arc pour en modifier la courbure et/ou connaître les mots associés et la force de la liaison.

Si l'espace d'un premier graphe est encombré ou peu rempli, recommencez en élévant ou abaissant le seuil pour réduire ou augmenter le nombre d'arc et de mots (en sollicitant le bouton THEME, puis le bouton GRAPHE). 

LES MOTS - Les mots en rouge sont des noeuds à grande fréquentation. 
Les noeuds en noir sont moins fréquentes et comptent moins de cinq liaisons. 
La taille des caractères est proportionnelle à l'intégration du mot dans le réseau.

LES TRAITS - Les tracés en rouge correspondent aux cooccurrences directes avec le mot-pôle (cercle des amis). 
Les tracés en bleu distinguent les relations que les amis du premier cercle ont entre eux.
Les tracés en noir intègrent le deuxième cercle (les amis des amis).
La force du trait (pointillé ou maigre en noir, maigre ou gras en couleur) correspond à la densité de la liaison.
n cas de recouvrement, draguer légèrement un mot pour le rendre visible. Maintenir la souris sur un point quelconque d'un arc pour en modifier la courbure et/ou connaître les mots associés et la force de la liaison.

LES MOTS - Les mots en rouge sont des noeuds à grande fréquentation.
LES TRACES - Les tracés en rouge correspondent aux cooccurrences directes avec le mot-pôle (cercle des amis).
LES traces en bleu distinguent les relations que les amis du premier cercle ont entre eux.
LES traces en noir intéressent le deuxième cercle (les amis des amis).
La force du trait (pointillé ou maigre en noir, maigre ou gras en couleur) correspond à la densité de la liaison.
Conclusions and perspectives

• Common: Academic discourse

• Differences
  • structure, style, morphosyntax, lexical

• Perspectives
  • Annotation
  • Refined selection of texts
Thank you for your attention!

Céline Poudat (celine.poudat@unice.fr)

Peter Follette (follette@unice.fr)