Evaluation of authorial presence in academic genres
Céline Poudat (1), Sylvain Loiseau (2)

(1) CORAL, Université d’Orléans,
278 rue Colverts, 45160 Olivet – France
[+33] (0)2 38 66 65 34
Celine.Poudat@univ-orleans.fr
(2) Modyco, Université Paris X Nanterre
1, rue du Roi de Rome, 92 500 Rueil Malmaison – France
[+33] (0)1 47 08 97 02
s.loiseau@mageos.com

Academic genres are highly subjected to social interactions that enable to regulate and stabilize text linguistics properties. Consequently, linguistic phenomena can be apprehended thanks to formal criteria. We propose here to demonstrate that different degrees of authorial presence can be evaluated with low-level linguistic criteria.

The hypothesis was tested on two French corpora representing three different genres. The first one is composed of 42 recent linguistic scientific articles and the second of 11 contemporary philosophy works, i.e. commentaries and essays. The texts were marked up with XML, which enabled to take genre structural and linguistic properties into account and to distinguish notes, citations or examples. Scientific articles and philosophical texts could then be considered without the elements that have their own characteristics (such as citations, examples or appendix). The texts were finally parsed and tagged thanks to the Cordial© parser.

The Personal Pronoun (PP) repartition on marked-up texts enables to obtain first indications, that have to be refined thanks to the study of the correlation between verbal tense distribution and PP. The verbal lemma lexicon was finally extracted (per tense and per person) in order to refine the results.

No VOUS and TU pronouns was found in any of the two corpora. Indeed, no interlocutive relationship is established between the authors and their readers, contrary to Anglo-Saxon research articles which are more reader-centred. Nevertheless, the high presence of ON (Scientific articles: 24,3% et Philosophical texts: 11%) et IL (S: 33,13% et P: 32%) pronouns reveals a more impersonal style. The present tense, which is the most used in scientific articles (80% of conjugated verbs in both corpora), is besides strongly correlated to impersonal pronouns. This provides articles with the impersonal and immaterial style recommended by tradition. The present tense coupled with 3d PP is correlated with the author withdrawal from his discourse and consequently, with an objectivity effort. Thus, the present tense is rather employed with positively-valued or non valued verbs.

The JE PP indicates a high responsibility of the author, and its significant presence (7,99 %) in scientific articles could indicate an adaptation of the European style to the Anglo-Saxon dialogical style. In philosophical discourse we observe a clear variation of the JE PP between essays (4 %), which are more subjected to author specificities and commentaries (1,6 %), in which the philosophical community point of view is more important. In both corpora, JE is the most employed PP in the future and conditional tenses. The authorial presence is more obvious when his discourse is prospective or hypothetical. Negatively valued verbs were besides correlated with the future tense in scientific articles. Indeed, the future tense is often employed to express the limits and the perspectives of the study (verbs such as se contenter/to content o.s. with or se limiter/to limit o.s., which were absent from the other tenses). A similar phenomenon was observed in philosophical discourse, as the lexicon was negatively valued in conditional and 1st person utterances. There is indeed a correlation between valuation modality (real or potential) and positivity. The NOUS PP is less representative of the author presence than the JE PP in research articles, as verbs conjugated
with JE are more subjective than with NOUS (verbs such as espérer/to hope or aimer/to like were scarcely employed with NOUS). The exclusive and inclusive NOUS types can besides be detected thanks to verb semantic values: venir/to come or voir/to see are generally coupled with a NOUS that includes the reader. In the same way, NOUS indicates a weakened responsibility in philosophical discourse, as well as an invocation of the addressee or of philosophical tradition to support the author’s interpretation. According to the verbal tenses and lemmas, a clear partition can be observed between inclusive and exclusive NOUS: the former NOUS is used with enunciative verbs whereas the latter is often negatively or imperfectively (nous nous efforcerons/we shall try to…) valued.

Taking genres into account is in both cases necessary to evaluate the author presence with low-level variables. The results were finally gathered within an evaluation grid that enabled to observe the authorial presence in the texts, according to the genre standards. We obtained more precise results for scientific articles, and they are particularly interesting for computer linguistics applications such as automatic summarizing, in which the identification of the authorial presence would enable to increase the relevance of the information selection.

The two corpora finally enabled to illustrate occidental academic discourse characteristics. The withdrawal and the discretion of both readers and authors allow to warrant the objectivity and immateriality that are specific to ‘true’ discourses whereas the author presence is correlated to metadiscursive verbs. Nevertheless, scientific and philosophic discourses reflect different levels of responsibility. Indeed scientific articles are very bureaucratic and the development of scientific multinational standards requires the progressive use of reader-friendly style. The importance of objectivity and the absence of individuality that characterize most of the continental academic writing is in conflict with the individualistic authorial self expected in English academic writing. On the contrary, philosophical discourse is less normative and the author presence is distributed among a higher number of pronouns, which may be due to the influence of literary rather than bureaucratic tradition and evolution.