

Labeling Discourse to Build Academic Persona

Geneviève Bordet
Université Paris Diderot, France

Abstract

Academic research is an increasingly competitive activity and scientific writers are under the constant pressure of getting published. Getting past the screening device of the scientific abstract is widely based on the ability to create a discourse that is perceived as coherent, considering the target discourse community and the communicative intention. This study focuses on the use of general 'labeling' nouns as a factor of coherence and rhetorical persuasion in scientific abstracts, with specific interest in terms that are determined by an anaphoric 'this'. Based on the study of PhD abstracts written in English by English and French applicants in several disciplines, my research aims to identify the factors of success and failure in the handling of this device by native and non-native writers. Labeling nouns are identified and semantically classified for each discipline, according to linguistic origin. Case studies show that success requires adequate lexical choice of labeling nouns. It is also based on an appropriate semantic and syntactic connection between the selected labeling noun and the segment it refers to, which requires sufficient general and scientific language proficiency. Didactic applications are then offered in order to raise scientific writers' awareness of the impact of this type of cohesive device on their credibility.

Introduction

This article focuses on the use of general 'labeling' nouns as a factor of coherence and rhetorical persuasion in scientific abstracts. Those engaged in scientific research find themselves under increasing pressure to publish in a highly competitive academic market. To do so, writers must persuade the target audience of the relevance of their research for the scientific community, most often within the limited space of abstracts which are more and more used as a screening device. While originality and relevance are assessed in the context of previous research in the field, the overall argumentative coherence is based on internal textual, lexicogrammatical and rhetorical devices which ensure that the account of a lengthy and often complex research process appears as a perfectly clear demonstration. Hence the paramount importance of research on cohesive devices in scientific discourse. Very influential research by Halliday and Hasan (1976) described several types of cohesive devices, including reference. One of these devices is based on the use of 'this' as a determiner followed by a 'general' or 'labeling' (Francis 1994) noun (ex: 'this strategy' 'this problem', etc.). It has attracted attention for various reasons. First, it appears as a powerful tool to connect successive steps of the argumentation as it establishes the link between a new focus of interest and a previous piece of information which it refers to, inside the text. At the same time, the selection of a noun re-categorizes the reference thereby offering both a cohesive link and an interpretation while using a minimum of words. Here is an example taken from a PhD thesis abstract, in the domain of didactics of mathematics:

Example 1:

'As of yet, there has not been research into teachers' knowledge of logarithms. This study was an attempt to fill **this gap**.'

Both the cohesive role of 'this' and the interpretive potential of 'general' (Halliday and Hasan 1976) or 'labeling' nouns (Francis 1994) have already been widely studied (Schmid 2000; Charles 2003; Mahlberg 2005; Swales 2005). Specific interest has focused on the possible ambiguities resulting from the use of 'unattended this' (Swales 2005). However, the choice of effective labeling nouns in the case of 'attended this' has attracted less study. The linguistic skills required for their handling in scientific writing has received even less attention. Nevertheless, all studies in that field converge towards the question of their contribution to the building of a coherent discourse.

The scientific abstract, as a highly condensed genre, requires strong conceptual coherence to attract the reader's attention while allowing clear interpretation despite the quasi absence of discourse and metatextual markers (Hyland 2004). It is therefore important to better understand how the lexicogrammatical patterns are intertwined so as to form the 'texture' (Halliday and Hasan 1976) of the text to efficiently guide the reader's attention from one step of argumentation to the next.

Based on the study of PhD abstracts written in English by native and non-native (French) writers in various disciplines, this article addresses the conditions under which a labeling noun determined by 'this' offers a contribution to the construction of a coherent account of the research, therefore enhancing the writer's scientific credibility (Charles 2003). The comparison should help with identifying potential disciplinary and linguistic specificities, and verifying the assumption that handling this cohesive device might be more difficult for non-native writers. The final section is a contribution to avenues opened up in didactics. It aims at enhancing writers' awareness of the impact of this type of device on the construction of effective argumentation.

1. 'This' + 'Labeling noun': building coherence upon a cohesive device

I give here a brief overview of the interaction between lexicogrammar, cohesion, coherence and genre before getting on to the specific cohesive roles of 'this' and 'general' or labeling nouns.

a. Cohesion, coherence and success of the genre

Constraints of modern research i.e. the pressure for publishing and the resulting inflation of available academic literature, have made clarity and conciseness cardinal values. Only easily understood abstracts have a chance to attract their reader to the more time-consuming reading of full papers or dissertations. The argumentation must appear as clearly structured and perfectly coherent to convince the reader, especially in as short a form as the abstract. While textual coherence is assessed 'with respect to the context of situation' (Halliday and Hasan 1996: 23), it is based on the use of internal cohesive devices. This phenomenon has been widely studied, among others, by Halliday and Hasan who define cohesion as follows: 'Cohesion occurs where the INTERPRETATION of some element in the discourse is dependent on that of another' (Halliday and Hasan 1976: 4, original emphasis).

While Halliday and Hasan emphasize the role of syntactic patterns, Hoey (1991: 2005) focuses on textualizing collocational patterns. Gledhill (1995), Hyland (2004) and Biber et al. (2007) study the interaction between lexicogrammatical cohesion and the rhetorico-semantic structure of the moves. They all show that each genre requires a specific type of coherence.

b. 'This' as a cohesive device

Referential determiners, because their interpretation requires contextual information, contribute to weaving a semantic chain. The determiner 'this' is also characterized by its plasticity, since it can point at extratextual reality (ex: 'in this research') or refer to previous or following items of the text. 'This' can be 'supported' or 'unsupported' (Swales 2005, Wülff, Römer and Swales 2012). 'Unsupported this' or 'this' used as a pro-form (ex: 'this shows that') always refers to a whole segment of text while 'supported this' can either determine a reiterated term ('a strategy..., this strategy') or a segment which is therefore re-categorized (see example 1). The simple reiteration of the term only reinforces cohesion. By contrast, when 'this' determines a labeling noun, it 'offers a higher-level recontextualization of the previous text: in other words, the writer can provide the reader with an interpretation of what he or she has just read' (Swales 2005: 3). As shown below, only specific types of terms can ensure this type of re-categorization.

c. Nouns as a cohesive and interpretive device

Halliday and Hasan (1976: 274) point at a specific class of "general nouns" which are "a borderline case between a lexical item (member of an open set) and a grammatical item (member of a closed set)". Whether they are termed as 'anaphoric nouns' (Francis 1986), 'carrier nouns' (Ivanic 1991), 'labeling nouns' (Francis 1994), 'signaling nouns' (Flowerdew J. 2003), 'shell nouns' (Schmid 2000), their common characteristic is precisely that they have no fixed meaning or denotation and can only be defined functionally (Schmid 2000). Their 'structure-inherent semantic gap' (Schmid 2000: 76) is precisely what makes them a privileged vehicle for the combination of a cohesive and an interpretive function. However, there are considerable differences between the above mentioned approaches. While Halliday and Hasan 1976: 275 define 'general nouns' as 'superordinate members of major lexical sets' including words such as 'man' or 'person', Francis, Flowerdew and Schmid focus on the interpretive and re-categorizing function of epistemological terms such as 'structure' or 'criterion'. Schmid (2000) insists on the functional role of 'shell nouns', which encapsulate complex chunks of information to connect them with another piece of information. Flowerdew (2003) focuses on a specific function of 'abstract words' which explicitly 'signal' the intended organization of the text. Francis (1994) also points at the organizational function of nominal groups called 'labels'. She shows that, beyond organizing the text and its progression, 'labels' can also signal the writer's evaluation of the propositions which they encapsulate. She mentions the fact that retrospective labels 'are almost always preceded by a specific deictic such as 'the', 'this', 'that' or 'such' (Francis 1994: 85). Charles (2003) also uses the term 'label' to highlight their contribution to the construction of PhD authors' stance in various disciplines. So does Yamasaki (2008) to emphasize their evaluative function. Since my aim is to understand how this type of term contributes to reinforcing the cohesion of an abstract while guiding the reader's interpretation of the proposed demonstration, I will from now use Francis's term 'labeling nouns' while the acronym 'LN' will designate 'labeling nouns' preceded by 'this'.

2. A CONTRASTIVE APPROACH

My approach to the cohesive and interpretive role of the labeling noun determined by 'this' is based on a study of PhD abstracts written in English by native and French speakers, in five disciplines. While PhD abstracts do not play the same decisive role in the 'gatekeeping' process as research article abstracts, they offer an interesting reflection of would-be 'insiders' (Hyland 2004: 78) representation of the disciplinary linguistic standards, in a genre characterized by its conciseness.

a. The corpus

The corpus studied is comprised of 150 PhD abstracts taken from five disciplines or specialized domains: didactics of mathematics (DM), materials science (MS), biology (BIO), semantic web (SW) and mathematics (Maths). It should be noted that all of the theses are

monographs, compiled these still being a minority in France. For each domain or discipline, 30 texts were collected: all the texts were written in English but half of them were written by native English speakers (L1), and half by French speakers (L2). So as to face the impossibility to verify each writer's native language, it was assumed that the validation of the dissertation by an institution in an English-speaking university was proof that the writers' proficiency in English qualified him or her to be considered as a native writer of English. The abstracts were downloaded from various Anglophone university websites and from the Proquest dissertations database (Proquest 2014) The sub-corpus of 75 texts written in English by French doctoral students was collected thanks to the French open archive database 'Hyperarchives en ligne' (Hal Archives Ouvertes 2014). An interesting characteristic of this database is that, for each PhD, an abstract written in French and in English is available. Therefore, while the studied corpus only includes texts written in English, for the French writers' English texts, a parallel corpus of abstracts written in French could be used in the future for comparison of the use of labeling nouns in native and non-native language. Table 1 provides quantitative data as to the number of words in each sub-corpus, considering disciplinary and language origin:

Table 1: word count for each discipline according to linguistic origin

Discipline	Native writer	Non-native writer
BIO	5980	4467
DM	7430	3780
Maths	4195	3312
MS	5230	4003
SW	5896	4111
Sub-total	28731	19673
Total	48404	

b. A five-step approach

The methodology includes five steps, from corpus to text analysis.

First, the abstracts are marked for moves using Bhatia (1993) and Hyland (2004)'s approaches of rhetorical moves in abstracts. For Bhatia, the rhetorical macrostructure of abstracts includes four moves: Introduction – Methods – Results – Conclusion. Hyland offers an augmented version of this typology: Introduction – Purpose – Method – Product – Conclusion. The rhetorical structure of the collected abstracts is confronted with these two models. The rhetorical structure of each of the collected abstracts is identified then compared with these two models.

Secondly, the occurrences of 'this' for each of the ten sub-corpora are identified and counted, using the concordancer Antconcs 'wordlist' and 'concordance' functions (Anthony 2014).

In a third step, labeling nouns determined by the anaphoric 'this' are identified through a manual analysis combined with the use of Antconcs's various functions. This third step comprises 3 phases: 1) all occurrences of 'this' followed by a determined term are identified using the 'concordance' function; 2) the 'file view' function allows an analysis at text level to select labeling anaphora 3) labeling anaphors are listed according to the authors' disciplinary and linguistic origins;

Based on previous research by Schmid (2000), Francis (1986; 1994) and Charles (2003) (see above), it was considered that only nouns preceded by 'this' which re-categorize a previous set of information could be considered as 'labels'. Therefore, cases where the noun preceded by 'this' reiterates a previous term (ex: 'a strategy.....,This strategy...') were excluded. In addition, such frequent textualizing items as 'this study', 'this dissertation', etc. (Vičič 2013) were also excluded. In a fourth step, labeling nouns are classified semantically for each sub-corpus.

Finally, the fifth step, based on case studies of texts marked for moves, sets the focus on the cohesive and interpretive functions of the labeling noun determined by 'this' and its influence on the argumentative progression.

3. RHETORICAL STRUCTURE, FREQUENCY AND SEMANTIC STATUS OF 'THIS'+LABELING NOUN IN PHD ABSTRACTS

The five steps described above aimed at assessing both the quantitative and qualitative role of the labeling noun determined by 'this' in PhD abstracts. This role was studied at corpus, sub-corpus and text level.

a. Rhetorical structure

So as to determine the extent to which the use of LN contributes to shifting the focus from one move to the next, each text was manually marked for moves, based on an analysis of the argumentative structure. While the study used Bhatia's four moves and Hyland's five moves' models, as mentioned in the description of our 5 step contrastive approach (cf. 2.b), the results give evidence of a slightly different model since a majority of the abstracts appear to be organized along the following four moves: contextualization, research proposal, method, results. As compared with Bathia's 'introduction' move, a distinction was introduced between contextualization and research proposal. It must be noted that while 'contextualization' nearly disappears in materials science, it sometimes takes up to a third of the text in didactics of mathematics or 'web semantics'. The most frequent rhetorical pattern in the studied corpus also differs from Hyland's model in so far as difficulties in separating 'product' and 'conclusion' led to keeping one single move for 'results'. Materials science abstracts, for instance, tend to privilege the description of results over their interpretation.

b. Frequency data at corpus and sub-corpus level

While occurrences of 'this' could be counted automatically, using the "word list" function of Antconc, the identification of labeling nouns required going back to the whole abstract, using the 'file view' function. Manual analysis allowed the retrieval first of 'this' used as a determiner, then of LN.

Example 2 taken from the 'semantic web' corpus:

'Previous works in ontology management, primarily for ontology mapping, ontology integration and ontology evolution, have exploited only one form or another of ontology management in restrictive settings. However, a distributed and heterogeneous environment makes it necessary for researchers in **this field** to consider ontology interoperability in order to achieve the vision of the Semantic Web.'

The term 'this field' re-categorizes 'previous works in ontology management' since it confers the status of a research domain to all the activities listed under 'ontology management'. It does not repeat any of the terms that have been used previously in the text.

Table 2 presents the collected data as to the frequency of occurrences of 'this', whether it be supported or unsupported, in each discipline and in L1 and L2, as compared with LN frequency:

Table 2: frequency of 'this' in the corpus: number of tokens and percentage of LN / 'this' occurrences

	Bio L1	Bio L2	Maths L1	Maths L2	SW L1	SW L2	DML1	DML2	MSL1	MSL2	Total
this	34	40	28	38	46	52	66	32	30	41	407
LN	15	16	3	11	12	12	8	5	3	11	96
% LN/this	44,11	40	10,7	28,94	26,08	23,07	12,12	15,62	10	26,8	23,74

Table 2 shows that 'this' is used in all sub-corpora. The proportion of 'this' used as the determiner of a labeling noun as compared with the overall occurrences of 'this' varies from 10% to 44%. The average percentage is 23%. This proportion seems relatively high, considering that 'this' can also be used to point at the dissertation (ex: 'this paper'), to reiterate a previous item, or as a pro-form ('unattended this'). Further studies based on a comparison between general and scientific language would be necessary to determine whether LN are more frequent in scientific than in general language and in abstracts.

Besides, table 2 gives evidence of major differences in the use of LN between disciplines and between linguistic origins. As seen from this limited corpus, the most productive disciplines are biology and, to a lesser extent, the study of 'semantic web'. Comparison between native and non-native writers comes into contradiction with the intuitive assumption that native writers might feel more comfortable with the handling of this rather sophisticated device. As an example, non-native writers in mathematics use LN nearly three times more than native writers. It is also the case in Materials Science. This discrepancy shows that while non-native writers might be assumed to be at a disadvantage in the mastering of LN, it does not stop them from using this device.

c. Semantic classification of labeling terms according to discipline and linguistic origin

To identify the conditions under which LN contributes to building coherence, focus was set on the lexical choice of labeling noun and the nature of its relation with the segment it refers to.

Therefore, and to better assess their semantic role, labeling nouns were listed and classified according to the type of semantic re-categorization they operate. The study of the listed nouns in context shows that three cases appear prototypical of the labeling nouns' semantic status.

First, a large proportion of these nouns are used to create a connection between a piece of information and conceptual categories which are emblematic of scientific research epistemological values ('hypothesis', 'model', 'system', etc...). They point at heuristic concepts which can be considered as ensuring communication at an ideational level (Halliday 1994). They will be referred to as representing 'epistemological values'.

Example 3: taken from the semantic web L1 corpus:

'The discovered user profiles, which are represented by the centroids of the Web user session clusters, are then used to make usage-based collaborative recommendation via a top-N weighted scoring scheme algorithm. In **this scheme**, the generated user profiles are learned from usage data.'

The term 'scheme' encapsulates the whole preceding segment and reclassifies the operations described as part of a general research strategy.

A second category includes the segment it refers to inside a 'super-ordinate' category of a specific disciplinary taxonomy (ex: in biology, 'repertoire', in mathematics 'expression'). These nouns refer to disciplinary concepts. They can be considered as operating at both an ideational level, by conveying specialized information, and an interpersonal level by signaling the writer's and the reader's belonging to a same discipline. These nouns will be referred to as 'super-ordinate' nouns.

Example 4 taken from the biology L2 corpus:

'Nevertheless, FYTTD1 seems to be involved in RNA processing. Since RNA could be another support for epigenetic information, the characterization of the exact role of **this protein** will be an exciting challenge for the future.'

'This protein' re-categorizes FYTTD1 as a protein, therefore including it in a conceptual category and taxonomy of the discipline.

A third category expresses the author's evaluation of the relevance of the information: the information referred to may for instance be described as a 'gap', a 'problem' or 'an incongruence'. This categorization operates at an interpersonal and attitudinal level (Halliday 1994). These terms will be referred to as 'attitudinal'.

Example 5 taken from the didactics of mathematics L1 corpus

'While the findings from the study show students reporting moderately high levels of self-efficacy, the perceived levels of self-efficacy in the different domains is not reflected in achievement in mathematics. The reasons for **this incongruence** are explored in the context of student achievement trends.'

'This incongruence' refers to a gap between 'perceived self-efficacy' and real degree of 'achievement'. By pointing at a problem, the use of LN opens perspectives for the research proposal.

The distribution of these three categories along the disciplines and linguistics origins has been studied, as shown in Fig. 1:

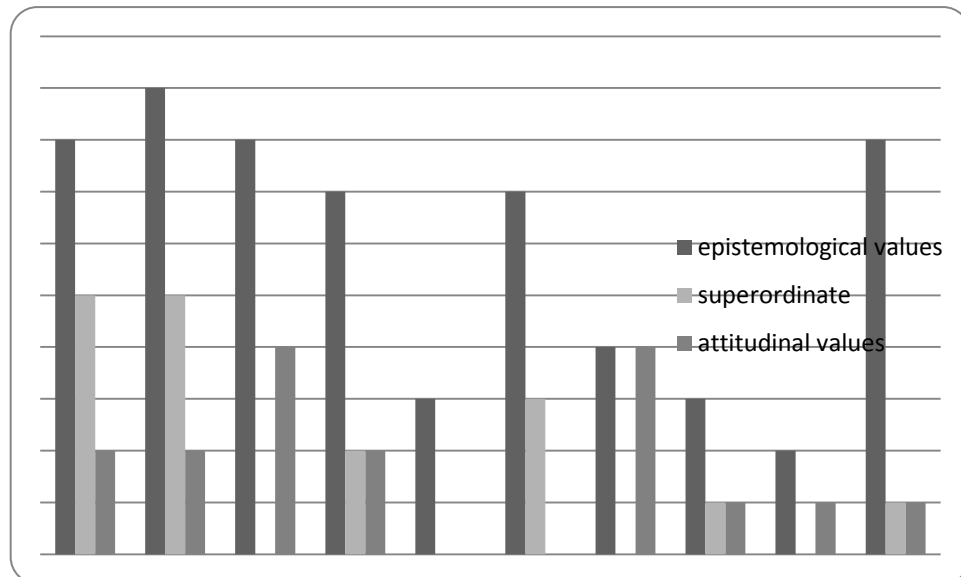


Figure 1 Distribution of LN according to semantic status

Results show that the first category ('epistemological value') is by far the most frequent and is distributed across all the disciplines.

The 'super-ordinate' terms nearly all belong to biology, which could be interpreted as a reflection of the importance of classifications in that discipline. It should be noted that 'problem' is used once as a 'super-ordinate' in the mathematics corpus, while other occurrences of this noun belong to the attitudinal category.

'Attitudinal' nouns are present to various extents in 8 of the 10 sub-corpora: no occurrence could be found in either mathematics L1 or L2 sub-corpora. The study of the corpus shows that 'attitudinal' terms mostly point at values connected with what Swales' (1990) CARS model calls 'establishing a niche', as shown by the recurrence of terms such as 'gap', 'problem', 'issue'. The emphasis is set on the relevance of the research proposal. One possible interpretation of their absence in mathematics is that it is a cumulative science, where research proposal is based on an extension rather than a discussion of previous work. The contrast is striking with the didactics of mathematics L1 corpus where research proposals are mostly based on a critical approach to the learning environment.

The semantic approach of the lexical selection of labeling nouns provides evidence that the 'semantic gap' inherent to labeling nouns is 'filled' by three types of values and concepts related to scientific research: epistemological concepts, super-ordinate concepts of a discipline and evaluation of the relevance of information. The distribution of these categories across the corpus varies according to the disciplines.

4. CASE STUDY: FACTORS OF SUCCESS OR FAILURE OF THE LABELING MECHANISM

Lack of space makes it impossible to present here the 96 cases of LN which have been retrieved and studied as described in the previous section. The five examples given above were all successful instantiations of re-categorization of a previous textual segment by a labeling noun. The extract given below is taken from an abstract written in English by a French author, focusing on the study of the 'semantic web'. This particular example was selected for two reasons. First it combines successful and unsuccessful uses of LN, which provides interesting insight into the factors of failure or success. Besides, the possibility of comparing the English abstract and the parallel French version might help to assess the extent to which inadequate choices are due to lack of mastery of a second language. The extract is marked for rhetorical moves which allows for an assessment of the LN's influence on the argumentative progression.

Example 6:

[Contextualization_beg] Since the 90s, the Internet has tremendously evolved in terms of number and diversity of available services. In **this trend**, proxies are playing a central role and are spread all over the net.(...)[Contextualization_end]

[Research proposal_beg] We propose to make proxies evolve from simple passive intermediaries to an open platform with advanced caching functionalities stressing on maintaining a high level of compatibility with existing paradigms. [Research proposal_end]

[Method_beg] We ground our developments on two qualities already available in proxies (...) To extend the proxy perceptivity upstream to the users, we propose a mechanism to maintain a session, based on proxy-cookies. Therefore, we propose two new HTTP directives similar to those used for cookies. **This paradigm** also allows us to install personalized services with the support of the interaction concept and user profile. **This last one** allows us to tackle with mobility problems and to install proxies independently of physical network infrastructures. (...)[Method_end].

This extract offers an exemplification of successive success and failure in building coherence, through ensuring the shift of focus from one move to the next, or inside a move. 'This trend' encapsulates the initial statement (underlined) and turns it into a 'given' fact. It ensures both coherence and persuasiveness by shifting the focus from 'the Internet' to 'proxies', and from contextualization to research proposal ('make proxies evolve'). In the 'method' move, 'this paradigm' encapsulates the first step of the described methodology (underlined), thereby providing a sound basis for the next step. It also expands the semantic status of 'mechanism' and 'directives' to show them as part of a more complex device, thus reinforcing the author's capacity to transfer local to global knowledge.

Conversely, 'this last one' fails at establishing a clear connection with either of these terms: 'support', 'interaction concept', or 'user profile'.

The French version might help us identify the reference:

'Ce paradigme nous permet également de mettre en place des services personnalisés en supportant le concept d'interaction et de profil utilisateur. **Ce dernier point...**'

Literally translated in English:

'This paradigm also allows us to install personalized services by supporting the concept of interaction and profile user. **This last point...**'

In French, the labeling noun 'point' clearly relates to 'the concept of', which was not the case in the English version, due to an erroneous combination of terms in the segment 'the interaction concept and user profile'. In addition, the choice of 'point' rather than 'one' is more effective to encapsulate the previous step. It appears that in English, a syntactic error combined with a wrong semantic choice of labeling noun blocks the flow of information, thus preventing the author from creating a connection between two steps of the 'method' move. This example shows that successful handling of LN requires the capacity to build a complex network of signals. While adequate semantic lexical correspondence prevails, a wrong syntactic construction may also stop the reader from identifying the connection between LN and the encapsulated information. Comparison with the French version shows that the chances for the author to convince the reader are hampered by poor handling of second language, in both aspects. This example is of course not sufficient to conclude to a systematic connection between the level of ability in handling labeling nouns and the linguistic origin of the author. However, it does show that while LN is an interesting device to guide the reader's attention along the argumentative steps, failure might severely impede the author's scientific authority by weakening the force of argumentation.

A second case of problematic use of LN in an abstract written in English by a French writer provides an opportunity to compare factors of failure with the previous case. This extract is taken from an abstract of the biology sub-corpus:

[Context_beg] (...) Today, few data are available about the pathophysiology of tularemia and about the virulence mechanisms of F. tularensis. Most of the studies have been conducted with F. novicida, a weak pathogen in humans. Moreover, F. tularensis study has been hampered by the lack of specific genetic tools. [Context_end] [Method_beg] In this work, we characterised in F. Philomiragia (....) Then, we developed a directed mutagenesis method of F. Tularensis (...).The respiratory burst-inhibiting AcpA was recognized as a virulence factor in F. novicida, data further extended to F. tularensis. To check **this hypothesis**, we developed a mouse model of infection by the respiratory route.

The textual segment 'this hypothesis' refers to is very difficult to identify for several reasons. One is that no previous information points at an interrogation or a claim. The second reason is that while the phrase 'to check this hypothesis' seems to express the research proposal, this interpretation is contradicted by the fact that it seems to point at one of the successive steps of a research process ('we characterized', 'then we developed') marked by the use of temporal connectors and the simple past. It may be assumed that the hypothesis to which the LN refers is that 'AcpA' is a 'virulence factor', which would be coherent with the research gap mentioned at the beginning of the extract: 'few data are available (...) about the virulence mechanism'. However, this interpretation comes in contradiction with the presentation of the existence of this virulence factor as a result of the research (was recognized as a virulence factor). The overall effect is confusing: not only does this ambiguity disrupt the flow of information but it also prevents the reader from clearly identifying the research proposal.

The comparison of this extract with the French version shows that the French and the English text are strictly parallel. Therefore, in this case, failure in the use of LN cannot be attributed to lack of mastery of English but rather to inadequate construction of scientific discourse.

5. PERSPECTIVES: DEVELOPING EAP LEARNERS' AWARENESS

These case studies highlight the importance of developing novice academic writers' awareness of the rhetorical impact of this type of referential device. Two types of activities are offered here: critical reading and making writing choices. Both aim at getting students to reflect on the connection between argumentation, and lexicogrammatical choices.

a. Critical reading

The objective of this activity is training students to identify the referential chain and assess its impact on cohesion and persuasion.

The extract given here is taken from the didactics of mathematics corpus:

Example 7:

[Contextualization_beg] Several research studies have confirmed that high-school and undergraduate students have a very poor knowledge of logarithms and logarithmic functions. One of the possible reasons for students' difficulties could be an insufficient teachers' knowledge of **this subject domain**. As of yet, there has not been research into teachers' knowledge of logarithms. [Contextualization_end]

[Research-proposal_beg] This study was an attempt to fill **this gap**. The deeper understanding of teachers' knowledge, particularly subject matter knowledge and related pedagogical skills, leads towards improvement of instructional approaches for more effective teacher training. [Research-proposal_end]

The first task is to identify the successive 'contextualization' and 'research proposal' moves.

The students then have to identify the two occurrences of LN ('this subject domain', 'this gap') and their respective referential segments (underlined).

Finally they are asked to assess the impact of these two LNs on the rhetorical structure, showing how 'this subject domain' clarifies the focus of the contextualization, while 'this gap' grounds the research proposal.

A variant could be to first offer the text with gaps to fill, corresponding to the two labeling nouns and study the consequences of lexical choices on coherence and persuasion. (Ex: replace 'this subject domain' by 'logarithms' or 'mathematics').

b. Experimenting writing choices

The objective of this activity is to raise students' awareness of the impact of their writing choices on the production of a coherent and persuasive abstract. Students are required to perform the following tasks:

- Write the abstract of your own research project (alternatively, write the abstract of a research paper), using the stereotypical move structure: contextualization, proposal, method, results.

NB: if you wish to modify this structure, explain how and why: which move would you add or eliminate?

- Use a labeling noun determined by 'this' to shift the focus from one argumentative move to the next.
- Mark the textual segment the labeling noun refers to.
- Try using a different labeling noun and observe the consequences on the shift of focus.

The point here is to encourage students to experiment with various lexicogrammatical alternatives while observing their influence on the argumentation. This type of exercise also promotes the idea that there is no 'correct' or 'incorrect' choice but more or less effective choices as to the production of successful argumentation.

Conclusion

The examples I have studied give evidence that the use of a 'labeling noun' determined by an anaphoric 'this' can be an efficient strategy to guide the reader's attention along the rhetorical moves. At the same time, it offers the reader an interpretation of an item of information which has been previously given in the text.

The success of this device is based on two main factors. First, the space made available by the selected nouns' 'structure-inherent semantic gap' (Schmid 2000: 76) creates the connection with general or disciplinary scientific values or concepts. Secondly, to successfully ensure the shift of focus from a piece of information to its re-categorization through generalization, the handling of LN requires the construction of an efficient network of semantic and syntactic connections. Failure to help the reader identify the encapsulated segment results in an interruption of the information flow. This interruption does not only undermine textual coherence: it also weakens the author's credibility insofar as he/she is unable to impose an interpretation.

Quantitative and qualitative comparison between abstracts produced by native and non-native writers give no evidence that the latter tend to use less LN than native writers. The range of terms is similar, so are the semantic categories the labeling nouns belong to. However, case studies of LN use, based on the comparison between the operation realized in second language (English) and native language (French), show that failure may be due to syntactic errors as well as incapacity to establish a correspondence between the labeling noun and adequate semantic signals in the textual segment it refers to. Even though these findings do not allow one to decide whether efficiency is related to language proficiency or to the writers' level of expertise in scientific writing, they still give evidence that all scientific writers' attention should be drawn to the potentialities and risks of basing one's argument on the use of labeling nouns.

References

- Anthony, Lawrence. Antconc: a freeware corpus analysis toolkit for concordancing and text analysis. Available at: <http://www.laurenceanthony.net/software/antconc/>
- Biber, D., Connor, U. and Upton, T.A. (2007) *Discourse on the Move: Using Corpus Analysis to Describe Discourse Structure*. Amsterdam/Philadelphia: John Benjamins.
- Bhatia, V.K. (1993) *Analysing Genre: Language Use in Professional Settings*. London: Longman
- Charles, M. (2003) 'This Mystery...': a Corpus-based Study of the Use of Nouns to Construct Stance in Theses from Two Contrasting Disciplines.' *Journal of English for Academic Purposes* 2 (4), 313–326
- Flowerdew, J. (2003) 'Signalling Nouns in Discourse.' *English for Specific Purposes* 22 (4), 329–346
- Francis, G. (1986). *Anaphoric Nouns. Discourse Analysis Monographs No. 11*. Birmingham: University of Birmingham
- Francis, G. (1994) 'Labelling Discourse: An Aspect of Nominal-group Lexical Cohesion'. in *Advances in Written Text Analysis*, ed. by Coulthard, M. London: Routledge, 83–101
- Gledhill, C. (1995) *Collocations in Science Writing*. Tübingen: Gunter Narr Verlag
- Halliday, M. A. K. (1994). *An Introduction to Functional Grammar*. London: E. Arnold
- Halliday, M. A. K. and Hasan, R. (1976) *Cohesion in English*. London: Longman
- Hoey, M. (1991) *Patterns of Lexis in Text. Describing English Language*. Oxford: Oxford University Press
- Hyland, K. (2004) *Disciplinary Discourses: Social Interactions in Academic Writing*. Ann Arbor: University of Michigan Press/ELT
- Hyper Archives en Ligne (Hal). Centre National de la Recherche Scientifique. Paris. Available at : <https://hal.archives-ouvertes.fr/>
- Ivanič, R. (1991). 'Nouns in Search of a Context: A Study of Nouns with Both Open-and Closed-System Characteristics'. *IRAL-International Review of Applied Linguistics in Language Teaching* 29 (2), 93–114
- Mahlberg, M.(2005) *English General Nouns: A Corpus Theoretical Approach*. Philadelphia/Amsterdam: John Benjamins Publishing
- Proquest Dissertations and Thesis Database. Proquest. Ann Harbor, Michigan (USA). Available at : <http://www.proquest.com/products-services/pqdt.html>
- Schmid, H.J. (2000) *English Abstract Nouns As Conceptual Shells: From Corpus to Cognition*. Berlin: Walter de Gruyter
- Swales, J. (1990) *Genre Analysis: English in Academic and Research Settings*. Cambridge applied linguistics series. Cambridge: Cambridge University Press
- Swales, J. (2005) 'Attended and Unattended 'this' in Academic Writing: A Long and Unfinished Story'. *ESP Malaysia* 11, 1–15

- Vičič, Š. G. (2013) 'The Difference a Word Can Show: A Diachronic Corpus-Based Study of the Demonstrative "this" in Tourism Research Article Abstracts'. in *Variation and Change in Spoken and Written Discourse: Perspectives from Corpus Linguistics*. Ed by Bamford, J., Cavaleri, S, and Diani, G. Amsterdam: John Benjamins, 223-238
- Wulff, S., Römer, U. and Swales, J. (2012). Attended/Unattended This In Academic Student Writing: Quantitative And Qualitative Perspectives. *Corpus Linguistics and Linguistic Theory* 8 (1), 129–157