

**AUTHORIAL DEVELOPMENT IN RESEARCH WRITING:  
CODING CHANGES IN GRAMMATICAL SUBJECT**  
**Desenvolvimento da Autoria na Escrita de Artigos de Pesquisa:  
Codificando Mudanças do Sujeito Gramatical**

Ann MONTEMAYOR-BORSINGER (Instituto de Lingüística, Universidad de Buenos Aires and Instituto Balseiro, Universidad Nacional de Cuyo)

**Abstract**

*This work presents a new taxonomy for coding changes in grammatical subjects in research writing, and applies it to study papers written over time by ten different scientists. For each scientist a first and two later publications have been selected, with a maximum time span between the first and the last article of sixteen years. The linguistic analysis of the texts is based upon a new classification of grammatical subject, which identifies a set of four main functions: Participant, Discourse, Conventional and Instantial Subject functions. The distinction between Conventional and Instantial Subjects is particularly relevant for the study of authorial development. Conventional Subjects are readily available wordings that are part of the terminology of a given research field, whereas Instantial ones are multifunctional nominal groups especially composed to fit a given stretch of discourse. Results indicate a correlation between a higher frequency of use of Instantial elements and writer expertise, showing that as writers gain experience, they become increasingly capable of manipulating language to reflect the complexity of their research activities.*

**Key-words:** *Research article writing, authorial development, discourse functions of grammatical subject, multifunctional nominal groups.*

**Resumo**

*Este trabalho apresenta uma nova taxonomia para codificar mudanças nos sujeitos gramaticais na escrita de artigos de pesquisa, aplicando-a a artigos escritos por dez cientistas diferentes, em diferentes momentos de suas carreiras. Para cada cientista, uma publicação no início da carreira e duas publicações posteriores foram selecionadas, com um*

*intervalo máximo de tempo entre o primeiro e o último artigo de dezesseis anos. A análise lingüística dos textos baseia-se em uma nova classificação de sujeito gramatical que identifica quatro funções principais: as funções de Sujeito Participante, do Discurso, Convencional e Instancial. A distinção entre Sujeito Convencional e Sujeito Instancial é particularmente relevante para o estudo do desenvolvimento da autoria. Os Sujeitos Convencionais são fraseados imediatamente disponíveis que fazem parte da terminologia de um dado campo de pesquisa, ao passo que os Sujeitos Instanciais são grupos nominais multifuncionais especialmente compostos para se adequarem a um dado trecho do discurso. Os resultados indicam uma correlação entre frequência maior de uso de elementos Instanciais e experiência do autor, mostrando que, à medida que os autores ganham experiência, tornam-se cada vez mais capazes de manipular a língua para que ela reflita a complexidade de suas atividades de pesquisa.*

**Palavras-chave:** *escrita de artigos de pesquisa; desenvolvimento da autoria; funções discursivas do sujeito gramatical; grupos nominais multifuncionais.*

## 1. Introduction

Over the years the study of writing has been recognised as being important in educational, professional, and research contexts, with a view of promoting more genre-based and effective support for specialised English instruction. The systematic study of article introductions by Swales (1981, 1990), and of experimental articles by Bazerman (1984,1988) has provided new insights in the organisation of information in scientific texts. Further research, central to the structuring of discourse in the work place, has been developed by Grabe and Kaplan (1989, 1996) from the perspective of contrastive rhetoric. Biber (1988) has explored textual variation by examining linguistic features that reflect functional dimensions of text structure, and in later work contrasts registers such as conversation with academic writing (Biber 1999).

Bazerman (1998) has pursued research interests in the direction of the constant creation of new concepts, a process that is being studied both in oral and in written institutional discourse. Iedema (1999) examines how meanings are increasingly being technologized in project planning. Beacco et al (2002) study new forms of discourse concerning transgenic plants that have brought about changes in the status of science. From a systemic-functional perspective, this unceasing evolution of scientific English has been of particular interest to Halliday (1998). He demonstrates that changes in scientific discourse are enabled by changes in grammatical resources and sees the creation of new language and concepts as a 'regrammatization' of experience by means of grammatical metaphor.

The present study uses a longitudinal approach to account for how professional scientists change the nature of their discourse as they gather experience by analysing the type of grammatical subjects chosen in research articles published over a maximum time span of sixteen years. It is different from other longitudinal studies, in the sense that it takes as its point of departure socially validated texts published in refereed journals. Previous existing longitudinal studies of academic texts have often taken as their point of departure essays written by students entering university, rather than published material. A well-known study is that by Berkenkotter, Huckin and Ackerman (1989) presenting an analysis of three different assignment introductions written by the same PhD student. By analysing these essays, Berkenkotter et al study how this student, who was not familiar with the conventional structure of articles, starts acquiring the genre knowledge characteristic of a research community. Haswell (2000) also analyses improvement in college writing by examining essays used by Washington State University to place students into composition courses. However, as the texts were not written in professional settings for an audience of peers, Haswell himself asks whether the changes found in such texts can truly constitute improvement in professional writing and development of expertise, an observation which points towards the relevance of using socially validated texts such as the published research articles of the present study.

Gosden (1993), working on published articles, emphasises the importance of choices in grammatical subject, because such choices

affect the way scientists structure interaction with their research community. Grammatical subjects in research articles have also been the focus of work by Tarone et al (1998) who look at the use of active and passive voice in astrophysics texts, while McKenna (1997) has classified subjects using the Davies (1988) and Gosden (1993) taxonomy to examine the writing up of facts in three engineering reports.

A thought-provoking taxonomy of grammatical subjects used in academic texts has been offered by MacDonald (1992). She distinguishes between Phenomenal Subjects that have to do with the researchers' object of study per se as in '*Shakespeare did x*', and 'knowledge making' Epistemic Subjects that have to do with the methods, conceptual tools and previous studies researchers bring to bear on that object of study as in '*The New Historicism is characterized by x*'<sup>1</sup>. In a recent longitudinal study of grammatical subjects, Hewings (2001) uses the MacDonald taxonomy to compare geography essays written by students in their first, and in their third year of academic study. She finds a greater proportion of these more 'Epistemic' Subjects in third year.

The present longitudinal study proposes a new 'Conventional'/ 'Instantial' distinction for wordings in Subject position. Conventional Subjects are commonly used wordings within the research field concerned, as in '*Chiral gauge theory is an interesting model*'<sup>2</sup>. In contrast, Instantial Subjects are expressions that have been especially composed to fit a given stretch of discourse, as in '*Whether a chiral gauge theory (CGT) with an arbitrary fermion content can be consistently quantized or not is still an open question*'. The need for a new taxonomy originated from coding difficulties encountered in previous research that examined authorial development in refereed journals (Montemayor-Borsinger, 1999, 2001, 2003). It was found in interviews with the authors of the articles that most of the highly specialised terms used in their papers are both 'Phenomenal' and 'Epistemic' in the sense given by MacDonald because they identify

<sup>1</sup> Both examples have been taken as such from MacDonald 1992, pp. 543 and 544.

<sup>2</sup> In all the examples grammatical subjects are in bold italics.

both objects of study and knowledge making elements that push science forward. However, these highly specialised terms differ in the level of writer creativity involved. The Conventional/Instantial distinction apprehends changes affecting grammatical subject as ten researchers gather experience in publishing articles in international journals of physics. The corpus is composed of texts directly ‘from the work place’ as it were, i.e. socially validated texts published in refereed journals for an audience of critical peers. It takes as its baseline the first published research article written by physicists who have already been apprenticed into their discourse community.

The present analysis is influenced by the systemic-functional approach associated with Halliday, where the clause is seen as a combination of three different strands of meaning, interpersonal meanings that organize clauses as exchanges, experiential meanings that organize clauses as representations of experience, and textual meanings that organize the interpersonal and experiential dimensions into strings of coherent messages. Within this approach, a study of grammatical subject is particularly relevant for exploring how these three different strands of meaning combine and map onto one another. Subject choice is crucial when designing sentences as the subject is “something by reference to which the proposition can be affirmed or denied” (Halliday, 1994:76). It specifies the element that is responsible for the success of the proposition, the element “on which the validity of the information is made to rest” (Halliday, 1994:76). In a dynamic perspective, as sentences unfold, each successive instance of subject choice pushes the text forward in particular ways that determine how this text is going to be interpreted.

The paper is organised as follows. The next section discusses how the corpus was selected. Section 3 considers the general research methodology, which is established around two main axes, one linguistic, involving the new subject taxonomy, and the other statistical, presenting the data analysis. Findings are presented and discussed in Section 4. Finally, results are considered in the light of the different combinations of interpersonal, experiential and textual meanings that may be placed in subject position according to writer expertise.

## **2. Selection of the corpus of research articles**

The corpus, which provides the basis for applying the new taxonomy to the analysis of changes in five thousand grammatical subjects over a maximum period of sixteen years, consists of 30 published research articles with an average length of around four thousand words per article.

### **2.1. Selection of the research articles**

A series of interviews were conducted on the basis of which ten physicists were asked to select three papers each: the first paper they had written, one of their last papers, and a paper in between. In further interviews the articles were discussed with authors to ensure that the writing had been mainly theirs, with no external help or at most very minor changes done by editors. Moreover, publication of the articles in international refereed journals was necessary to make sure they were socially validated scientific texts both as regards language and scientific ideas. The scientists concerned were not necessarily native speakers, but all of them did many years of postgraduate studies and postdoctoral research in international institutions where English is the lingua franca, and work in different countries. They were asked to select the articles themselves because the majority has well over forty publications, sometimes co-authored. Moreover, the fact that ten different active members of the physics community, rather than one linguist, selected three papers each, increases the likelihood of the corpus of research articles being chosen at random and not concealing some hidden bias.

### **2.2. Journals where the research articles were published**

Physics is a highly internationalized branch of research, with several hundreds of journals. There exists detailed information in the *Journal Citation Report* about their relative importance, and in particular, about the ranking sorted to them by Impact Parameter. This parameter is the ratio between the number of published articles in a given journal and the number of citations referring to these articles, both within a

period of two years. For instance, if a journal has an Impact Parameter of 2, this means that its articles are cited twice in average in other ranked international journals within two years following publication. The higher the Impact Parameter, the more frequently cited are the articles from a given journal, with less than 20 journals having an Impact Parameter higher than 3.

The distribution of the articles from the present corpus as a function of the Impact Parameter centers around high Impact Parameter journals of nearly 3, indicating that the articles tend to come from relatively higher ranked and more prestigious journals. Articles from the corpus were published for instance in *Physical Review Letter* and *Physical Review D* from the American Physical Society, *Journal of Physics B* from the Institute of Physics, and in publications by North Holland-Elsevier Science and World Scientific Publishing Company.

### 3. Research method

The research method is established around two main axes, one linguistic, and the other statistical. To conduct a longitudinal survey focusing on subject changes, it was necessary on the one hand to set up the linguistic notions of ‘Conventional’ and ‘Instancial’ wordings to capture these changes and, on the other, to interpret the findings in reliable ways with the assistance of statistics.

#### 3.1. Design of the longitudinal study

The present study is longitudinal in the sense that the corpus was set up by taking publications written by the same respondents at different time intervals. Because the present research seeks to capture changes in the use of grammatical subjects in published articles, it was necessary to consider intervals of time running into years between first and last papers. A serious problem affecting longitudinal studies is that they tend to suffer case losses, the more so in the case of extended time spans. In order to prevent such loss of information, the present analysis was designed as a retrospective longitudinal study where there was one

data collection point. This data collection point took place two years ago, when each researcher was asked to furnish three articles published at different times.

### 3.2. Design of the taxonomy

Another important task was the design of coding frames required for the analysis of subject changes. Each category within a coding frame should be as unambiguous and non-overlapping as possible. The corpus has more than five thousand grammatical subjects, all of which were coded as belonging to one of four different classes: the Participant, Discourse, Conventional and Instantial classes. The Participant and Discourse classes, taken from Davies (1988) and Gosden, (1993) are easier to distinguish linguistically and are presented first. Then, the two new Conventional and Instantial classes set up in Montemayor-Borsinger (2002) will be discussed with examples from the present corpus. As explained in the introduction, these new classes had to be set up in view of coding difficulties encountered when using previous subject taxonomies.

#### 3.2.1. The Participant Class

Linguistic items signalling this class are personal pronouns or nouns concerned with researchers and their work such as *We* and *Our approach*, where authors appear openly in the text, albeit in different degrees. In the present corpus, subjects of the Participant Class are mostly worded as a direct *we* rather than the more oblique *our...* Examples of Participant Subjects are the following:

*Here we are mainly interested in the simultaneous diagonalization of (x) and (y). We neglect, for the time being, the exchange interaction in (z) to make clear the predictions of the one-electron picture ...*

*We used two samples for our experiments ...*

*Our initial analysis was extended to stable and metastable compounds of the following groups ...*

### 3.2.2. The Discourse Class

Linguistic items signalling this class are terms that belong to the highly conventionalised discourse of researchers naming parts of their work. It is a lexical set that includes expressions such as *Table x*, *Figure x*, *This section*, *The present paper* etc... which refer to the text and its parts and concern the discourse acts of reporting and discussing. The extract below shows in bold examples of subjects from the Discourse Class:

***The paper is organised as follows: Section II presents the experimental set up for the forthcoming experiments ...***

### 3.2.3. The new Conventional Class

The Conventional Class is realised by elements, mostly of a taxonomising type, that refer to entities and events belonging to experiments and theory within the realm of physics. These elements belong to the specialised language that is commonly used in science, and typically do not contain interactive elements.

In the Conventional Class, subjects are realised by nouns or 'of-type' nominal groups characterised by:

- absence of post-modification except 'of-type' nominal groups. Following Sinclair (1991), 'of-type' nominal groups are not seen as introducing prepositional phrases that function as qualifiers, but rather as introducing a second noun as a potential headword, or as forming double-headed nominal groups. For instance, when meanings are expressed with double-headed nominal groups, neither noun seems to be more significant or dominant, and to express these meanings the 'of' structure tends to require both nouns. Examples of these typical Conventional Subjects that are nouns on their own or 'of-type' nominal groups are shown in the context of their respective clauses:

*Under a magnetic field  $H$  **the compound** undergoes a transition to a ferromagnetic state, at very low temperatures and at normal pressure...*

***The field of semiconductor microstructures** has also profited from this technique...*

***The inverse of  $D$**  as usual can be calculated by a perturbation expansion...*

***The program** optimises the parameters that are set free to vary...*

***The ratio  $dH/dE$**  is appreciably changed for both types of fermions...*

- optional pre-modification by items such as deictics, numeratives and classifiers as shown by this, first and second-order and reaction-diffusion in the following examples:

*Finally, this photoassisted oxygen ordering might help to understand the differences observed between the short and long-term illumination experiments...*

***The first and second-order Born approximations** present a range of agreement with experiments compatible with the condition  $Zi/v$ ...*

*Finally, **reaction-diffusion equations** have been obtained for the macroscopic density of a system undergoing reaction processes, in which particles are created or destroyed...*

- optional pre-modification by adjectives describing an objective property of the phenomenon in question as shown by irreversibility and generalized in the following:

***The irreversibility line (IL)** has been observed in  $YBaZCu3O7$ , bulk [2], single crystals [3] and thin films [4] as well as in other high- $T$  superconductors...*

*Besides this technical difference, **the generalized Coulomb potentials** are similar for both geometries...*

However, if the adjective expresses the scientist's attitude or stance towards the entity, as for 'new and interesting' in an example

such as *A new and interesting feature of the high-ir, superconductors*, it will not be included in the Conventional class, but in the Instantial class discussed in what follows.

#### 3.2.4. The new Instantial Class

The term ‘instantial’ was inspired by Halliday (1998). It is used here in his sense of wordings especially created for the immediate requirements of reasoning within a particular stretch of discourse. In a similar way to Conventional Subjects, Instantial Subjects are realised by elements that refer to entities and events belonging to experiments and theory within a given research field. However, the difference is that these subjects have been especially formulated to create new combinations of meanings, whereas Conventional Subjects, as indicated by their name, are commonly used wordings within given research fields.

Instantial Subjects include at least one of the following components:

- pre-modification by interactive adjectives expressing writer stance
- post-modification by prepositional phrases or relative clauses functioning as qualifiers
- the nouns themselves contain lexical or contextual clues pointing towards interactive meanings expressing writer stance.

These more highly crafted elements are needed, for instance, to ‘package’ information in resourceful and innovative ways in the subject slot. To package information and express new wordings, the researcher may need to form complex nominal groups containing embedded clauses and phrases. Instantial Subjects are also used for issues that may not yet be established, and may be concerned with interpretation or controversy, in which case authors resort both to modification and to interactive elements. Alternatively, Instantial Subjects are used by writers once they have absorbed and made their own the substance with which they are working. In all these cases there is authorial presence, either because authors have modified subjects in such an extensive way that they no

longer belong to the purely taxonomic jargon of their area of research, or because authors added new strands of meaning indicating writer stance. In contrast, we saw above that Conventional Subjects identify elements that are taken for granted and which are already established. They are commonly used terms that have not been specially created, but rather belong to the taxonomic system or specialised language of the research field concerned.

Table 1 shows ten examples taken from the corpus with Instantial Subjects in the context of their respective clauses.

	INSTANTIAL SUBJECTS
1	<i>The magnetic susceptibility that in other intermediate valence compounds shows a maximum at some finite temperature increases monotonically with decreasing temperature, until order sets in.</i>
2	<i>In addition, inelastic light scattering experiments for light propagating along the planes of the structure are possible.</i>
3	<i>Whether a chiral gauge theory (CGT) with an arbitrary fermion content can be consistently quantized or not is still an open question.</i>
4	<i>The failure in achieving a satisfactory representation of the selected data leads to various kinds of modifications of the previous decisions.</i>
5	<i>In fact, the strongest evidence for the universality of the gravitational interaction involves electrons, protons and neutrons.</i>
6	<i>A new and interesting feature of the high-ir superconductors was clearly demonstrated by the experiments of Müller et al. [1] in ceramic LaBaCuO.</i>
7	<i>Attempts to free the results from the selection of distorting potentials may not be pursued through a neglect of the effects brought by long-range forces on the transition amplitudes.</i>
8	<i>For our problem, the equations corresponding to the eigenstates with total spin*** take the form: ... [Equation (24)]</i>
9	<i>Joining together eqs. (4.13), (4.14) and (4.15), the only coefficients that survive are ** related by *Equation (4.16)*.</i>
10	<i>The reaction-diffusion equations obtained from the asymptotic expansion coincide with those proposed ad hoc at the macroscopic level to describe the evolution of the density.</i>

Table 1: Examples of Instantial Subjects (in bold italics) taken from the data

I shall now discuss in detail each one of these examples of Instantial Subjects to illustrate how they are identified on the basis of the criteria presented above. A lower case letter indicates that the subject is preceded by other sentence-initial elements, as shown in the clauses in Table 1.

Example 1. *The magnetic susceptibility that in other intermediate valence compounds shows a maximum at some finite temperature* - The wording *magnetic susceptibility* on its own would have been classified as Conventional. It has been post-modified by the researcher with a defining relative clause especially worded in order to give an adequate evaluation frame for what follows in the sentence.

Example 2. *inelastic light scattering experiments for light propagating along the planes of the structure* - This subject is similar to the previous one. *Inelastic light scattering experiments* on its own would have been classified as a Conventional Subject. However, it now belongs to the Instantial Class because of extensive post-modification that packs information and helps optimise discourse flow.

Example 3. *Whether a chiral gauge theory (CGT) with an arbitrary fermion content can be consistently quantized or not* - Here the researcher has cleverly built in his stance towards a problem by introducing a question in nominalized form in the subject slot, reinforced by the expression 'is still an open question' which finishes off the sentence.

Example 4. *The failure in achieving a satisfactory representation of the selected data* - This is an interesting example of stance showing up not only in the noun *failure*, but also being reinforced by post-modification, especially by the word *satisfactory*.

Example 5. *the strongest evidence for the universality of the gravitational interaction* - Here *evidence* is pre-modified by the epithet *strongest* manifesting the scientist's position, giving this subject from the very beginning a strong Instantial flavour which is then strengthened by extensive post-modification.

Example 6. *A new and interesting feature of the high-ir superconductors* - This is another good example of pre-modification by the adjectives *new and interesting* expressing the scientist's attitude vis-à-vis *a feature of the high-ir superconductors*. The latter 'of-type' noun on its own would have been classified as a Conventional Subject.

Example 7. *Attempts to free the results from the selection of distorting potentials* - This subject is similar to Example 4 in that stance is conveyed not only from the choice of the noun *Attempts*, but also from the wording of the post-modification *to free the results*. A question arises concerning the adjective *distorting* in *distorting potentials*: is it an objective property of *potentials*, or does it express the researcher's subjective attitude towards *potentials*? In the present case it actually belongs to the set of commonly used terms within Atomic and Molecular Physics, the field of research of the paper concerned. Nevertheless, the choice of the word *Attempts* followed by extensive post-modification places this subject definitely in the Instantial class.

Example 8. *the equations corresponding to the eigenstates with total spin\*\*\**. The wording *equations* is highly conventional, but here again extensive post-modification indicates clearly that this subject belongs to the Instantial Phenomena category.

Example 9. *the only coefficients that survive* - In this subject *only* is a clue for including this subject in the Instantial category as the researcher has chosen to define a very particular subset of *coefficients*. Moreover, added post-modification worded as *that survive* is an additional indication of it being coded as Instantial.

Example 10. *The reaction-diffusion equations obtained from the asymptotic expansion* - Here again, *reaction-diffusion equations* on its own would have been coded as Conventional. However, the post-modification *obtained from the asymptotic expansion* indicating covert authorial presence gives this subject an Instantial character.

### 3.3. Design of the statistical study

The results obtained when coding subjects from the corpus according to the four classes examined above will now be discussed. The main question centres upon whether there are perceivable changes towards the expression of more complex Instantial Subjects, especially fashioned to create new wordings and involving added strands of interpersonal meanings.

A detailed discussion of the differences among the ten individual physicists is the topic of a forthcoming paper. However, Subject percentages for just one of the ten researchers are presented to help clarify how the whole set of results was obtained. Researcher R works in Theoretical Physics, in research topics related to Condensed Matter. The three papers selected by Researcher R were published in 1985 in *Physics Letters*, in 1990 in *Physical Review B*, and in 1999 again in *Physical Review B*, with fourteen years between Paper 1 and Paper 3. The time lapse between the first and the second paper is shorter (five years) than between the second and the third (nine years). Table 2 shows the relative distribution of subjects in the three papers of Researcher R.

Paper	Year		Participant Subjects	Discourse Subjects	Conventional Subjects	Instantial Subjects	Total Subjects
1	1985	0	19%	-	56%	25%	100%
2	1990	5	14%	-	57%	29%	100%
3	1999	14	13%	7%	47%	33%	100%

**Table 2: Distribution of subjects in the three papers of Researcher R**

Subjects in the Participant Class decrease from Paper 1 to Paper 2, and then the percentage stays virtually the same in Papers 2 and 3. The percentage of Conventional Subjects stays virtually the same in the first two papers, and then diminishes in the last. Parallel to this decrease over time in both Participant and Conventional Subjects, there is an increase in Instantial Subjects as Researcher R gains experience. This increase of Instantial Subjects, where authors can combine more subtle interpersonal strands with experiential meanings, compensates both for

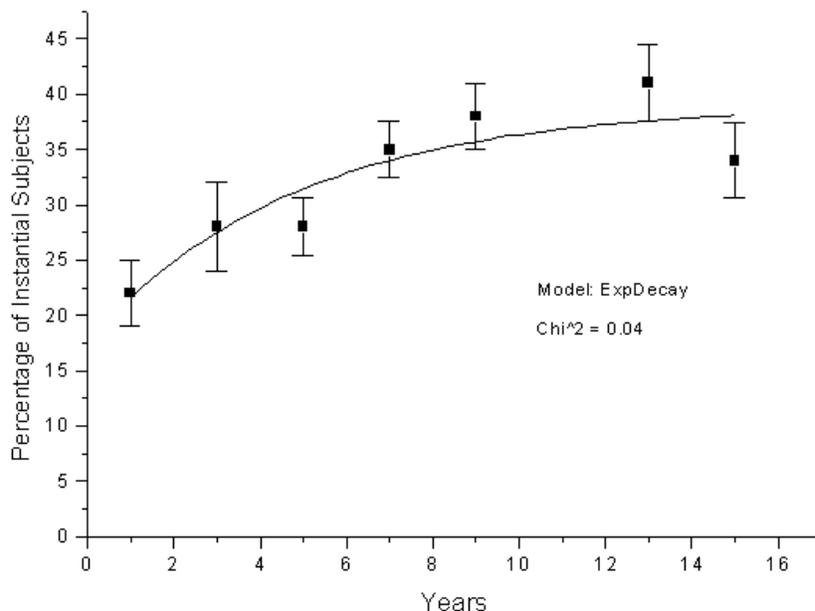
the decline in interactive Participant Subjects and for the decline of the essentially experiential Conventional Subjects. Researcher R's choice of subjects in the Discourse Class does not show a distinguishable trend over time. In Papers 1 and 2 no use is made of Discourse Subjects, and then in Paper 3 their percentage jumps to 7% - one of the highest percentages in the present corpus. An explanation is that these Discourse Subjects refer to five extremely complex figures in Paper 3. Paper 1 has no figures, and Paper 2 has only three very simple ones that were not referred to as such in the text.

Similar tables were created for each of the other nine researchers. Here the whole set of results for the ten researchers is considered as a unique sample to offer a more comprehensive view of the ways subject selection changes as researchers gain experience.

#### **4. Discussion of results**

##### **4.1. Trends for Instantial Subjects**

To help further clarify how the results were obtained, Figure 1 shows details of the trend in the use of Instantial Subjects with time. The set of data from the analysis of the thirty papers by all ten writers was ordered in function of time, with  $t=0$  adopted as a convention for the time of researchers' first publication, and spanning over a maximum of sixteen years. The figure shows seven points instead of eight over the maximum time span of sixteen years because between Year 10 and Year 12 there were no publications. The points represent the mean value of the data for all ten researchers within a two-year interval, and the bars centred on each point represent the standard deviation corresponding to this particular mean. The standard deviation measures the scattering of the data with respect to the corresponding mean. By taking the mean of the data obtained over two-year intervals, individual variations are moderated and it becomes easier to visualise more general and relevant trends. However, in order not to lose information concerning individual variations around each one of these means, Figure 1 also shows the standard deviation around each mean in the form of bars.



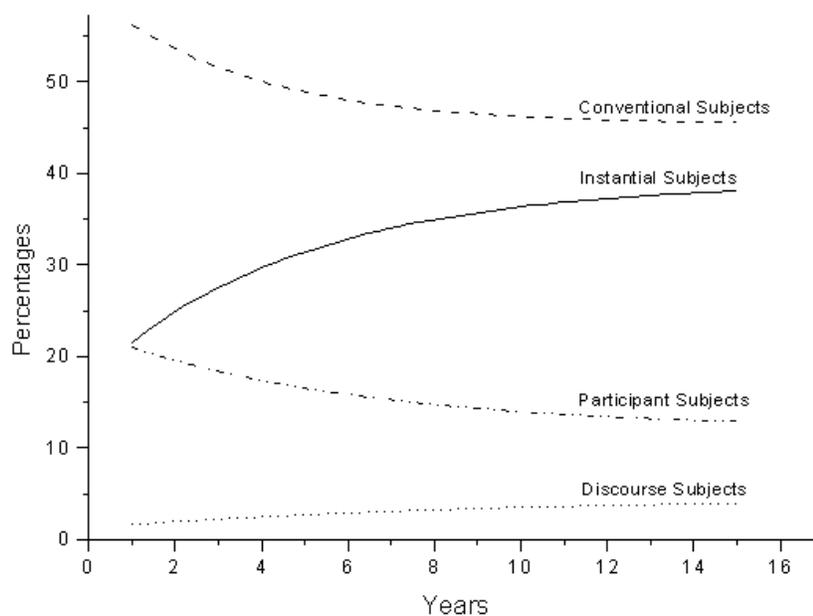
**Figure 1: Trends for Instantial Subjects**

An exponential curve fits the present set of data well, as is qualitatively shown by the figure itself and quantitatively stated by the value of the Chi-square parameter,  $\chi^2 = 0.04$ . For the present set of data, statistical tables for the Chi-square distribution show that values smaller than 0.6 indicate a confidence level for the curve near 100% (see for instance Freund and Wilson 1993).

#### 4.2. General trends

In a similar way to Instantial Subjects shown above, general trends were also identified for Conventional and Participant Subjects, and all three are now shown together in Figure 2. The trend for Conventional Subjects is shown by a dashed curve, the one for Instantial Subjects is shown by a full curve, and the one for Participant Subjects

is shown by a dotted curve. Results for Discourse Subjects indicated that they counted very little in the overall percentage of subjects, and did not show any definite trend. For the sake of completeness the curve for Discourse Subjects is reproduced on scale at the bottom of Figure 2.



**Figure 2: General trends for grammatical subject as writer expertise increases**

Figure 2 shows that Conventional Subjects in general represent the most numerous category, and represent more than half the subjects in first papers. Instantial and Participant Subjects both start off having roughly the same percentage, around 20%. These last two categories start moving in opposite directions and very quickly take on different values, while Conventional Subjects also change and decrease, but in a less spectacular way. Six years after a first publication, slightly fewer than half the subjects are selected from the Conventional category, Instantial Subjects are around 30%, and Participant Subjects tend to

decrease to around 15%. Fourteen years after a first publication, subjects in the Conventional Class tend to stabilise around 45%, Instantial Subjects around 36%, and Participant Subjects just above 10%. These trends are especially significant for the Conventional and Instantial Categories. In both cases the confidence level of the fitting is nearly 100%. In the case of the Participant Class, the confidence level goes down to 50%.

The general trends described above suggest that overall, experienced researchers make subject choices from the Conventional and Instantial categories, with comparatively fewer Participant Subjects. The order for latter years goes from Conventional (highest percentage), to Instantial and finally to Participant (lowest percentage). The picture for first papers looks less tidy: Conventional Subjects have the highest percentage, but then Instantial and Participant Subjects could start off in either order if we now bear in mind standard deviations. Standard deviation is particularly high for Participant Subjects for first publications, and is more than 13 for the first year. This means that although the general trend signals the use of Participant Subjects for a first paper as being around 20%, it could be, in some individual cases, nearer 30%, in which case Participant Subjects would become the second most important type of subject after Conventional ones, with Instantial Subjects in third place.

Over time, standard deviations diminish noticeably for Participant and Conventional Subjects, and stay around 6 for Instantial Subjects. A decrease in individual variations over time could be an indication of expert writing tending towards a greater perception of certain generic trends, which simultaneously would involve being able to compose Instantial Subjects when needed.

To sum up, data trends suggest the following:

- Conventional Subjects - obligatory - readily available choice at the onset of publishing research articles in science because they are part of the jargon, as it were, of the field of research concerned. Over time, Conventional Subjects remain the category with the highest percentage although they diminish in relative terms.

- Instantial Subjects - obligatory - more difficult subjects to manage effectively. As time goes by and as researchers become more experienced, their use increases.
- Participant Subjects - optional - a readily available choice at the onset of publishing research articles in physics. Over time, trends are for Participant Subjects to diminish by half, although a statistical analysis points towards a rather low confidence level for this result, which seems to be much more affected by individual choice especially for first articles.

#### 4.3. Conventional and Participant Subjects

Results point towards Conventional and Participant Subjects being easier to use than Instantial Subjects. Papers written by less experienced researchers show a greater proportion of these two types of subjects especially in first papers where the percentage of Instantial Subjects is still low. An explanation for their higher standard deviations could be that initially researchers focus their choices on one or another of these two types of subject depending on individual preferences. Researchers may choose to organise their texts mainly around Conventional Subjects, such as in the following extract from the corpus:

*'The **Boltzmann equation** determines the evolution of the distribution function for a gas of particles interacting through binary collisions... However, because of its complicated mathematical structure, **the Boltzmann equation** is hard to solve. In this sense, **simplified models** have become relevant in the study of relaxation processes.'* (Conventional Subjects shown in bold italics)

Alternately, researchers may choose to rely on Participant Subjects as shown in this other extract:

*'In this paper **we** consider the analog of the QED vacuum polarization for a lattice regularization of  $SU(N)$  gauge theories, using Wilson and Susskind fermions. **We** define our action over*

*an asymmetric rectangular lattice with different lattice spacings for spatial ( $a$ ) and temporal ( $a_t$ ) links.* (Participant Subjects shown in bold italics)

Both types of text organisation are possible options. If researchers choose a comparatively higher proportion of Participant Subjects, in the present corpus mostly the pronoun *we*, it does not necessarily indicate they have decided to take open responsibility for their work. Rather, especially in first papers, some researchers might choose to appear because they find it easier to arrange their writing by stating *In this paper we consider...* and *We define ...* in the manner some narratives are constructed. On the other hand, if researchers choose to use a comparatively higher proportion of Conventional Subjects, as their name of course indicates, are readily available options that can be preferred over Participant as they center the flow of ideas on the phenomena being discussed, for instance on *'the Boltzmann equation'*.

Results indicate that Participant Subjects are of an optional character, whereas Conventional Subjects are obligatory. This optional character of Participant Subjects has already surfaced in the analysis of individual cases discussed elsewhere (Montemayor-Borsinger, 2002), where some writers managed to make choices in such a way that Participant Subjects were virtually non-existent.

In contrast, it would be impossible for authors to do away with Conventional Subjects because of the very nature of research articles in science. In these texts Conventional Subjects are fundamental due to their essentially experiential nature: as discussed by Halliday (1998), what the grammar does in its experiential dimension is to set up theories of experience. Conventional Subjects, together with Instantial ones, construe a universe of things and relations and impose categories on scientists' perceptions of phenomena. A hypothetical option would be to express all the more obligatory experiential meanings concerned with scientific representation by Instantial wordings. However, this would prove far too cumbersome for the general flow of discourse, because of the added strands of meaning and extensive pre and post-modification present in Instantial Subjects.

## 5. Conclusions

If we now look at general trends in the light of the three strands of meanings distinguished by Halliday, results suggest that there would be a slight tendency for experiential meanings in subject position to increase: although Conventional Subjects decrease by 10% over time, Instantial Subjects tend to increase by 14%. As a reminder, experiential meanings are the representations of what is going on – in our present case the discussion of physical phenomena. Both Conventional and Instantial Subjects are experiential in nature. The difference is in the complexity of Instantial Subjects, which by means of pre and post-modification combines experiential meanings with interpersonal ones.

Regarding trends for interpersonal meanings, where writers manifest stance and interact with readers, results suggest a shift from overtly interactive Participant Subjects to more subtle and complex Instantial Subjects. In general, research has shown that interpersonal meanings are much more ‘moveable’ to different parts of the sentence and tend to be scattered throughout units of text, as discussed in detail in Halliday (1994:68-105 and 190-191) and in Hunston and Thompson (2000).

Moreover, further accounts have also shown how interpersonal meanings can adapt to different structures to such an extent that they appear as being ‘parasitic’ on other structural elements (Thompson 1996:65). This type of parasitism could be accounted for by the rise in Instantial Subjects that offer the possibility of interweaving experiential meanings with interpersonal ones. The capacity of finding optimum ways of combining experiential and interpersonal meanings, and enabling them textually by placing these elements in subject position, would be an important means of distinguishing expert from novice research writing.

This type of inquiry could be potentially useful for pedagogical applications. An important incentive for studying the ways in which more experienced writers of research articles manage grammatical subject is to help novice researchers enter more successfully into the publishing ‘arena’. In the author’s experience with academic writing workshops, young researchers often want to know how their published

work compares with that of leaders in their field, not only regarding results per se, but also regarding ways of presenting these results. Researchers publishing their first papers are acutely aware of the importance of mastering optimal writing strategies in a highly competitive publishing world. Rather than just seeking advice at the editing level, there comes a point when they want to discuss more detailed composing processes.

A greater focus on key elements such as grammatical subject can be a very effective way of helping, especially when time is short and the pressures to publish are great. The type of analysis presented here highlights possible options offered by the subject slot to suit different communicative aims and to enhance effective discourse flow. In particular, it is claimed that devising Instantial Subjects would be an important step towards giving a more “expert” tone to research articles, where writer choices are especially strategic as they affect the way in which findings are perceived by a research community. Text-based activities may be developed for apprentice researchers to see how these subjects are used in articles written by experts. An interesting outcome for practitioners of English is that a dialogue may naturally be established with specialists from other research communities, whose help is often crucial in selecting and analysing relevant texts from leading authors in their fields.

Another interesting outcome of designing such practices is that applied linguistics can become an important support for the teaching of language. These practices show ways of bridging the gap between theoretical linguistic input and practical methodology in language teaching. The present study has hopefully indicated that certain types of subject complexity can be analysed, and, by the same token, that they may be taught and learnt.

### **Acknowledgements**

I would like to thank Flo Davies for many enlightening discussions. Any shortcomings are my own.

Recebido em: 05/2004. Aceito em: 10/2004.

## References

- BAZERMAN, C. 1984 Modern evolution of the experimental report in physics: spectroscopic articles. *Physical Review*, 1893-1980 in *Social Studies of Science*, **14**: 163-196.
- \_\_\_\_\_. 1988 *Shaping written knowledge: the genre and activity of the experimental article in science*. University of Wisconsin Press.
- \_\_\_\_\_. 1998 Emerging perspectives on the many dimensions of scientific discourse. IN: J.R. MARTIN & R. VEEL (eds.) *Reading science: critical and functional perspectives on discourses of science*. Routledge.
- BEACCO, J.C, CLAUDEL, C., DOURY, M., PETIT, G. & REBOUL-TOURÉ, S. 2002 Science in media and social discourse: new channels of communication, new linguistic forms. *Discourse Studies*, **4.3**: 277-300.
- BERKENKOTTER C., HUCKIN, T.N. & ACKERMAN, J. 1989 Social contexts and socially constructed texts: the initiation of a graduate student into a writing research community. Technical Report **33**, National Center for the Study of Writing.
- BIBER, D. 1988 *Variations across speech and writing*. Cambridge University Press.
- \_\_\_\_\_. 1999 A register perspective on grammar and discourse: variability in the form and use of English complement clauses. *Discourse Studies*, **1.2**: 131-150.
- DAVIES, F. 1988 Reading between the lines: thematic choice as a device for presenting written viewpoint in academic discourse. *the ESPecialist*, **9.1/2**: 173-200.
- FREUND, R.J. & WILSON, W.J. 1993 *Statistical methods*. Academic Press.
- GOSDEN, H. 1993 Discourse functions of subject in scientific research articles. *Applied Linguistics*, **14.1**: 56-75.
- GRABE, W. & KAPLAN, R. 1989 Writing in a second language: contrastive rhetoric. IN: D. JOHNSON & D. ROEN (eds.) *Richness in writing*. Longman.
- \_\_\_\_\_. 1996 *Theory and practice of writing*. Applied Linguistics and Language Study, Longman.
- HALLIDAY, M.A.K. 1985, 1994 *An introduction to functional grammar*. Edward Arnold Publisher.

- \_\_\_\_\_. 1998 Things and relations. IN: J.R. MARTIN & R. VEEL (eds.) *Reading science: critical and functional perspectives on discourses of science*. Routledge.
- HASWELL, R. 2000 Documenting improvement in college writing: a longitudinal approach. *Written Communication*, **17.3**: 307-352.
- HEWINGS, A. & HEWINGS, M. 2001 Approaches to the disciplinary variation in academic writing: implications for syllabus design. IN: D. HALL & A. HEWINGS (eds.) *Innovation in English language teaching*. Routledge.
- HUNSTON, S. & THOMPSON, G. (eds.) 2000 *Evaluation in text: authorial stance and the construction of discourse*. Oxford University Press.
- IEDEMA, R. 1999 Formalizing organizational meaning. *Discourse & Society*, **10.1**: 49-65.
- MACDONALD, S. P. 1992 A method for analyzing sentence-level differences in disciplinary knowledge making. *Written Communication*, **9**: 533-569.
- MCKENNA, B. 1997 How engineers write: an empirical study of engineering report writing. *Applied Linguistics*, **18**: 189-211.
- MONTEMAYOR-BORSINGER, A. 1999 Research articles in science: a comparative study of linguistic choices. IN: H. BOOL & P. LUFORD (eds.) *Academic standards and expectations: the role of English for Academic Purposes*. Nottingham University Press.
- \_\_\_\_\_. 2001 Linguistic choices in two research articles in physics: study of an author's development. *the ESPecialist*, **22.1**: 51-74.
- \_\_\_\_\_. 2002 *Academic writing in the sciences: a focus on the development of writing skills*. Unpublished PhD dissertation, University of Glasgow.
- \_\_\_\_\_. 2003 A comparison of thematic options in novice and expert research writing. *ELA* **37**, Year 21: 37-52.
- SINCLAIR, J. 1991 *Corpus, concordance, collocation*. Oxford University Press.
- SWALES, J. 1981 Aspects of article introductions. *Aston Research Report No. 1*. University of Aston.
- \_\_\_\_\_. 1990 *Genre analysis: English in academic and research settings*. Cambridge University Press.

TARONE, E., DWYER, S., GILLETTE, S. & ICKE, V. 1998 On the use of the passive and active voice in astrophysics journal papers: with extensions to other languages and other fields. *English for Specific Purposes*, **17.1**: 133-136.

THOMPSON, G. 1996 *Introducing functional grammar*. Arnold.

*Ann Montemayor-Borsinger holds an MEd (Bristol) and a PhD in Linguistics (Glasgow). Her main activities involve teaching and research in Systemic Functional Linguistics for the Discourse Analysis master programme at the University of Buenos Aires and in Academic Writing at a centre for physicists and engineers. Her research interests also include applications of SFL to Spanish. [aborsinger@yahoo.co.uk](mailto:aborsinger@yahoo.co.uk)*