

*The research article and the science popularization article: a probabilistic functional grammar perspective on direct discourse representation*



*Discourse Studies*  
 Copyright © 2006  
 SAGE Publications.  
 (London, Thousand Oaks,  
 CA and New Delhi)  
 www.sagepublications.com  
 Vol 8(5): 627–646.  
 10.1177/1461445606064833

JANAINA MINELLI DE OLIVEIRA AND  
 ADRIANA SILVINA PAGANO

UNIVERSIDADE FEDERAL DE MINAS GERAIS – BELO HORIZONTE

**ABSTRACT** This article discusses the results of an investigation on discourse representation in a corpus of 34 million words constituted by texts in Brazilian Portuguese from two different genres: the research article and the science popularization article. Drawing on a systemic functional grammar perspective of language and pursuing a probabilistic approach, it focuses on the realization of *lexicogrammatical* systems of direct discourse representation as enacting interpersonal and social relationships. It is argued that the citation practices employed by writers in the genres studied negotiate discursive distance and are influenced by the *status* culturally and historically ascribed to scientific knowledge.

**KEY WORDS:** *discursive distance negotiation, discourse representation, genre, science*

## 1. Introduction

It is nowadays increasingly accepted that what is usually referred to as *scientific register* or *discourse* is actually what Myers (2003: 267), drawing on Fairclough (1992), describes as 'a terrain of competing discourses and practices', in which numerous participants play different roles in the making of science through specialist articles, textbooks, reports for lay readers and others (Halliday, 2004; Myers, 2003). Investigations developed by Halliday (1998), Halliday and Martin (1993), Swales (1990, 2004), to quote just a few, have been able to demonstrate that the research article and the language of science present rhetorical, grammatical and stylistic resources for organizing the general semiotic system so as to enact scientific communities' particular ways to conceive reality and knowledge production. Myers (1990) states that scientific articles, on the one hand, and popularization articles, on the other hand, present contrasting views of science, both contributing to maintaining the cultural authority of scientific knowledge.

Through their analysis of metadiscourse in texts aimed at the professional scientific community and texts geared towards readers of popular scientific articles, Crismore and Farnsworth (1990) contribute to demonstrate 'the groundlessness of the myth that views professional biological writing as consisting only of impersonal, factual statements devoid of such metadiscourse' (p. 118). White (2003) develops an intersubjective stance on the presence of different voices in texts, drawing on Bakhtin/Volosinov's dialogic perspective. He offers a typology of the ways writers construe texts as heteroglossically acknowledging the existence of alternative positions to their own or as monoglossically assuming the textual voice and the audience to share a certain amount of knowledge, beliefs and values which present the construed utterances as not risky or doubtful.

The research article and the science popularization article are currently acknowledged as two distinct genres. Calsamiglia et al. (2001), Orlandi (2001) and Zamboni (2001) have strongly made the point that the popularization article is not a simplified version of the research article, but a discursive reconstruction of scientific knowledge to an audience other than the academic one. Taking the research article and the science popularization article as distinct genres implies considering them as different vehicles of discursive practice. As Calsamiglia and Ferrero (2003: 147) put it, we need to explore 'the different settings in which knowledge circulates, setting out from the supposition that science forms part of the practices of human communities'.

In this article, we explore the different settings in which knowledge circulates by observing the ways writers not only signal, but accent the presence of different voices in scientific texts and in texts about science especially through devices linked to manifest intertextuality. We adopt here a useful convention suggested by Thompson and Ye (1991), and call *writer*, the quoting voice ultimately responsible for a text, and *author* the quoted voice whose discourse is represented. According to Fairclough (1992), we can expect a great variation in the degree of texts' heterogeneity and also 'in the extent to which their heterogeneous elements are integrated, and so in the extent to which their heterogeneity is evident on the surface of the text' (Fairclough, 1992: 104). The author goes on saying that every text has a potential group of other texts and voices it could call upon. This framework allows us to consider the *textualization* of scientific knowledge, in academic settings, as well as its *recontextualization*, in all other settings, as establishing social practices which interconnect different genres in an intertextual potential. Not only do written conventional sources constitute this intertextual potential, but also oral interviews between journalists and specialists and between specialists and the subjects who eventually take part in their researches, information drawn from the Internet or commonsense knowledge.

By adopting such a perspective we want to emphasize that although the research article and the popularization article are not directly connected in terms of their production, circulation and consumption, and of the actual intertextual

chains they establish with other specimens of the same genre, both the research article and the popularization article imbricate complex potential intertextual processes, thus problematizing the commonly held view that the former is merely a simplified version of the latter. By 'potential', we mean that some of the sources cited will be more frequently referred to, others more typical of a genre, others particularly used to achieve a rhetorical effect. In other words, even not participating in the same intertextual chain, the research article and the popularization article bear intertextual connections, ascribable to both manifest and interdiscursive intertextuality. Fairclough (1992: 128) states that 'particular genres are associated to "specific modes of (manifest) intertextuality"'. From the point of view of systemic-functional grammar and its dominant paradigmatic axis, which emphasizes every choice as meaningful, we could say that particular modes of manifest intertextuality instantiate in texts particular interpersonal and social relations between, on the one hand, writers and the authors they quote, and on the other hand, between writers and their readers.

The aim of this article is to analyze the ways in which the research article and the popularization article are built upon the presence of different voices in texts, looking for similarities and idiosyncrasies characteristic of each genre in the representation of direct discourse. We depart from Halliday and Matthiessen's (2004: 29) contention that 'while construing, language is always also enacting: enacting our social relationships with other people around us', to investigate the social relationships enacted by particular modes of signaling the presence of different voices in academic texts and in texts about science. In this article, we address the following questions: how frequent is direct discourse representation in the research article and in the popularization article? In what ways are the grammatical systems of direct projection in these distinct genres organized differently? In what ways are they similar? How do the systems of direct projection in the genres observed relate to the *status* culturally and historically ascribed to scientific knowledge?

We adopt a theoretical stance on direct discourse representation taken from the researches developed by Caldas-Coulthard (1994) and Calsamiglia and Ferrero (2001). Caldas-Coulthard (1994: 297) states that reported talk is a cleaned up version of real talk, so 'the representation of speech is a simplification and a reduction of the characteristics of real interaction'. Calsamiglia and Ferrero (2001) explain that every quote, regardless of the discursive genre imperative for literality, suffers a double intervention: first, the quoted discourse is removed from its context of occurrence, and second, manipulated so as to fit the context of reception, according to the writer's communicative purposes. So, direct discourse representation is here taken, both in the research article and in the popularization article, not as a resource through which adherence can be claimed to an original discourse, but as a rhetorical resource that, playing the game of manifest intertextuality, allows for identifying the quote as partially reproducing some original discourse.

## 2. *Theoretical assumptions*

As previously stated, the theoretical framework we adopt here draws on a type of grammar which is systemic and functional. It is systemic because 'language is represented in the form of system networks, not as an inventory of structures' (Halliday and Matthiessen, 2004: 23) and it is functional because these system networks are seen as a resource for producing meaning. Therefore, although structure is an essential part of a semiotic phenomenon description, priority is given to the meanings which arise from patterns of choice made within these structural systems. Texts are seen as the intersection point between two axes: one that is structural or syntagmatic and organizes a language in terms of 'what goes together with what', and another axis which is systemic or paradigmatic and organizes a language in terms of what 'could go instead of what'. In this sense, 'a text is the product of ongoing selection in a very large network of systems' (Halliday and Matthiessen, 2004: 23).

When we look at a text and its particular structural configuration, we are in fact trying to find out what kind of interface the choices made on the paradigmatic axis, instantiated in that particular text, have with the world and the social processes we engage in. The basic functions of a language can be said to be: 'making sense of our experience, and acting out our social relationships' (Halliday and Matthiessen, 2004: 29). Functional grammar systematizes these functions in three metafunctions: i) the ideational metafunction, which through its logical and experiential components, construes representations of the human experience; ii) the interpersonal metafunction, which construes language as interactive and personal; and iii) the textual metafunction, which discursively organizes the previous two metafunctions in cohesive and continuous moves. Each metafunction is organized in systems whose elements play a role in the organic configuration of the whole system, that is, of language itself.

Texts are related to the system through instantiation. When different texts share patterns of instantiation, from the system perspective, they may be called a register, that is, a functional variety of language (Halliday and Mathiessen, 2004). In this article, we focus on a macrotextual category other than register – genre – to explore its role in discursive and social practices. In order to reconcile different views on this category, we use genre on the one hand as an operational concept to probe into our corpus based on Genre Analysis and Critical Discourse Analysis accounts (Fairclough, 1992; Swales, 1990, 2004) and, on the other hand, as a macrotextual category akin to register based on systemic-functional theory. In this sense, drawing on Halliday and Mathiessen's observations for genre, we ascribe to genre the possibility of being taken as 'a syndrome of lexicogrammatical probabilities' (Halliday, 2005: 85). The lexis and the grammar of a language should be understood as poles in a continuum, the former offering open ended options of choices of instantiation, the later offering more limited options. When observing genres, that is, syndromes of probabilities, we need to take into account that 'part of the meaning of choosing any term is the

probability with which that term is chosen' (Halliday, 2005: 65) and that is why lexical and grammatical choices may mean different things in different genres, where probabilistic profiles may be found to vary.

As Halliday (2005: 95) puts it, 'in a systemic grammar, the entire grammar is represented as a network of systems, each of which potentially being a candidate for quantitative analysis'. In this article, we are interested in grammatical systems that play a part in explicitly signaling the presence of different voices in texts, that is, when a clause functions as a direct linguistic representation of another linguistic representation. Systemic-functional grammar labels this metaphenomenon a projection, more specifically, a quote or direct speech. In systemic-functional grammar, while reports (indirect speech) are projected through hypotaxis, quotes are projected through parataxis, when the projecting clause and the projected one have the same *status*, that is, they are independent. Direct speech or discourse is said to be the simplest form of projection: 'the projecting clause is a verbal clause, one of saying, and the projected clause represents what is said' (Halliday and Matthiessen, 2004: 445). Though considerably more restricted, the combination of a mental process with quoting is also found possible.

We believe the functional grammar treatment of direct discourse representation can be supplemented, on the one hand, through the adoption of a more discursive approach, and, on the other hand, through corpus investigations that have proved to be sensitive to naturally occurring data. Studies such as those developed by Swales (1985), though not adopting systemic functional grammar, have proved useful to the understanding of rhetorical motivations for the quoting behavior of researchers. Thompson and Ye (1991) have specially contributed to the categorization of reporting verbs and to the perception of the evaluative space their use opens up to the text writer particularly in academic texts. Thompson (1996a, 1996b) scrutinized a diversity of structural contexts a quote may occur in and of functions it may serve. Hyland (1999) presented an account of different quoting behavior patterns in terms of form and frequency in research articles written in different disciplinary communities motivated by cognitive styles and epistemological beliefs. Leech and Short (1981) and Semino and Short (2004) identified a significant number of intermediate possibilities of discourse representation between direct and indirect speech, thus allowing for speaking of a continuum. Biber et al. (1999) elaborated the first grammar based on electronic corpus and presented frequency data about the occurrence and position of reporting clauses in fiction and news texts. Caldas-Coulthard (1994) and Calsamiglia and Ferrero (2003) demonstrated the relevance of observing who gets quoted and who remains silenced.

Data from the corpus in Brazilian Portuguese we investigate in this research confirms the existence of a continuum of discourse representation, revealing numerous possibilities of signaling (or not) the presence of different voices in texts. Any account of these possibilities, as detailed as it may be, will always be partial. It seems necessary to recognize, on the one hand, that approaches which

consider direct and indirect speech two unproblematic categories are not sufficient to deal with the complexity of discourse representation, and on the other hand, that it is not possible to describe each and all varieties of the phenomenon. For the present discussion, we would like to problematize the proliferation of categories for discourse representation that may not be contributing to a better understanding of the kind of interpersonal and social cultural meaning they instantiate. In other words, we would like to contend that it is necessary to look at the continuum of discourse representation and try to find out what kind of interface the choices made on this paradigmatic axis have with the world and the social processes the texts under investigation engage in.

In order to contribute to the above discussion, we present here a contrastive study of the probabilistic grammatical profiles of the systems used to signal the presence of different voices in texts belonging to two genres: the research article and the science popularization article. We will then select a layer or area of the paradigmatic continuum of discourse representation and work on its syntagmatic realization in research articles and science popularization articles in order to better understand how the systems of direct projection in the genres observed relate to the *status* culturally and historically ascribed to scientific knowledge. By doing this we assume we are, at the same time, recognizing the complexity of the phenomenon of discourse representation and the need to examine the different interpersonal meanings realized in it.

### 3. *Study aims*

Our hypothesis for the interpretation of the interpersonal meaning of the structure in which a reporting verb is used to project a direct discourse representation is that there is increasing discursive distance between the writer and the quoted voice. Such a distance seems to indicate, on the one hand, the existence of asymmetrical authority relations between the quoting and the quoted voices, and, on the other hand, the interpretative orientation the writer prefers her readers to develop regarding the quoted discourse. The area of the discourse representation continuum we have chosen to observe in this article is that in which it seems there is the highest degree of separation between the voices of text writers and quoted authors. In order to better explain this point let us examine the following examples of direct discourse representation without reporting verbs. All examples are drawn from the corpus we investigated compiled from electronic issues of the general interest magazine *Isto É*, the science popularization magazine *Galileu*, the newspaper *Folha de São Paulo* and the *SciELO* database of research articles, all sources being Brazilian. A translation into English done by us is provided after each example. The corpora are presented in detail in the next section.

- 1) Maristela conta que para afastarem a solidão, muitas mulheres de oficiais do CLA estão adotando crianças da região. **'Há pouco, uma garota de 16 anos deu à luz**

**e já entregou o bebê para os pais adotivos.** A gravidez precoce é a norma por aqui. Muitas adolescentes têm filhos aos 12 anos. [Subcorpus *Isto É*]

1) Maristela tells us that so as not to feel lonely, the officers' wives of the CLA are adopting children from the nearby community. **'Recently, a 16 year old girl gave birth to a child and immediately afterwards gave her baby to its adopting parents.'** Teenage pregnancy is the norm around here. Many teenagers have babies at the age of 12. [*Isto É* subcorpus]

2) **Cunha (1993: 13) alerta para este fato:** 'Parece-nos que pronomes de primeira e de segunda pessoa tornam-se indeterminados na medida em que assumem a capacidade de englobar a "não-pessoa", o que pode acontecer inclusive com "eu".' [Subcorpus Ciências Humanas]

2) **Cunha (1993: 13) warns about this fact:** 'It seems to us first and second person pronouns become indeterminate as they assume a capacity to embody the "non-person", which may also happen with "I".' [Human Sciences subcorpus]

3) **Para o autor,** 'a noção de língua minoritária tomou forma através da análise de situações onde a distinção entre maioria e minoria era precisa.' [Subcorpus Lingüística]

3) **For the author,** 'the notion of minority language has arisen out of the analysis of situations where the distinction between a majority and a minority was necessary.' [Linguistics subcorpus]

In example 1, we notice there is an apparent fusion between the quoting voice and the quoted one. Apart from the quotation marks, which identify the clause complex in bold type as reproducing some previous discourse – presumably an oral interview –, there is no other signal of authority or evaluative distance. In example 2, we have a kind of direct discourse representation Thompson (1996a) calls discontinuous report. Notice the clause in bold type does not actually project the second one. The only typographic connection between the discourse represented as a locution between quotation marks and the clause in bold type is a punctuation mark (colon). In example 3, the report adjunct in bold type, as Thompson (1996a) calls it, introduces the direct discourse representation structure and again there is no reporting verb or projection process.

Just as direct discourse representation may occur without reporting verbs, reporting verbs may also occur in lexicogrammatical configurations other than direct discourse representation, as in the examples below:

4) **Os autores CONCLUEM** que o marcador é útil no diagnóstico diferencial de massas da adrenal, mas não tem relação com a agressividade biológica da neoplasia maligna. [Subcorpus Ciências da Saúde]

4) **The authors CONCLUDE** that the marker is useful for the differential diagnosis of adrenal mass, though its use does not allow for determining the biological aggressiveness of malign neoplasias. [Health Sciences subcorpus]

5) Muitas pessoas **ENFATIZARAM** a natureza objetiva dos dados sobre citação, mas também **OBSERVARAM** a sua unilateralidade: ser um membro da universidade exige muito mais do que pesquisa científica como a que é publicada em revistas do JCR. [Subcorpus Ciências Sociais]

5) Many people **EMPHASIZED** the objective nature of quotation data, but also **OBSERVED** its unilateral aspect: being a member of a university involves further demands other than doing scientific research such as that reported in JCR publications. [Social sciences subcorpus]

Example 4 is what we typically call indirect speech. The projecting clause in bold type operates in hypotaxis, through a reporting verb here signaled in capital letters and a dependent clause as metadiscourse. Example 5 may be seen as problematic because, although the verbs signaled in capital letters may be used as reporting verbs, and they actually involve linguistic representations of other linguistic representations, the grammatical relationship between the clauses in the complex is not that of projection, but of expansion. The conjunction group underlined in the example extends the previous clause by the addition of the information in the second one. Thus, though '*emfatizaram*' ('emphasized') is a verb which may be taken as functioning as a reporting verb, example 5 is not a clear-cut case to easily classify the whole structure as a report. In both examples, as there are no quotation marks, we can expect the writers are presenting discourses they read or heard in other contexts in their own words. Apart from the verbs marked in capital letters there is no other signal of the presence of different voices in discourse, but the writer's own voice.

The area of the discourse representation continuum we analyze in this article is that in which a reporting verb in a projecting clause projects as metadiscourse a locution, presented as reproducing a clause complex or a clause from some other piece of discourse. In a different way from what takes places in the examples above, we believe such a structure explicitly constructs the presence of asymmetrical authority and power relations between the quoting and the quoted voices and signals through particular lexical choices of reporting verb the interpretative orientation the writer prefers her readers to take regarding the quoted discourse. Under this particular case, we have found occurrences like the following:

6) **Após afirmar que informação não é opinião, ACRESCENTA:** 'Informação tem características intrínsecas, como responsabilidade . . .' [Subcorpus Ciências Sociais]

6) **After stating that information is not the same as opinion, [he] ADDS:** 'Information has intrinsic characteristics, such as responsibility . . .' [Social sciences subcorpus]

7) **Simões (1989: 40) REFERIU** que 'as verdadeiras e reais estimativas das taxas de mortalidade . . .' [Subcorpus Ciências Biológicas]

7) **Simões (1989: 40) SAID** that 'the true and actual estimates of mortality rates . . .' [Biology subcorpus]

8) 'Nossas escolas públicas em particular', **ele APONTA**, 'têm sido ótimos instrumentos de assimilação . . .' [Subcorpus Ciências Humanas]

8) 'Our public schools in particular', **he POINTS OUT**, 'have been great instruments of assimilation' [Human Sciences subcorpus]

9) '... acabam absorvendo o carbono que foi jogado na atmosfera com a queimada', **DIZ Volker Kirchoff, diretor do Instituto Nacional de Pesquisas Espaciais.** [Subcorpus *Isto É*]

9) '... end up absorbing the carbon that was thrown in the atmosphere with the burns', **SAYS Volker Kirchoff, director of the National Space Institute of Research.** [*Isto É* subcorpus]

10) Algo 'fascinante', **COMENTA Derrida** e 'insuportável', ocorre aí: (. . .) é exatamente . . . [Subcorpus Ciências Humanas]

10) Something 'fascinating', **COMMENTS Derrida** and 'unbearable', takes place here: (. . .) it is precisely . . . [Human Sciences subcorpus]

In examples 6–9, we have reporting verbs, in capital letters, in a projecting clause in bold type. These projecting clauses project as metadiscourse clause complexes or a clause taken from other pieces of discourse. Example 10, however, reproduces only partially the quoted discourse, that is, the portion of discourse represented as identical to the original discourse does not constitute an independent finite clause, but integrates the syntax of the quoting discourse. Occurrences like these were also selected in this research because they do signal, through the reporting verb and the quotation marks, not only the presence of different voices in discourse, but the existence of discursive distance and evaluative space.

Having defined the area of the continuum of discourse representation we are investigating and the hypothesis about the interpersonal meaning it instantiates, we proceed to select the systems which will allow the comparison of the frequency profiles of the genres here studied: the research article and the science popularization article. Halliday (2005: 96) states that 'once something has been shown to be a grammatical system, it is interesting ipso facto'; however, in order to examine experiential and interpersonal meanings, we need to focus on the systems of the clause. As our first task we took three systems: primary tense, Theme and logical interdependency. We then began to differentiate the occurrences selected for research. The projecting clauses, through their reporting verbs, were then classified according to present tense, like example 11 below, or past tense, like example 12. Only two percent of the occurrences were not in the present or past tense, as in example 13, and were thus disregarded so that the tense system could be treated in binary terms:

11) 'É bom esperar pelo menos duas horas antes de deitar', **DIZ** Jaime Eisig, gastroenterologista do Hospital das Clínicas, em São Paulo. [Subcorpus *Galileu*]

11) 'It's advisable to wait at least for two hours before going to bed', **SAYS** Jaime Eisig, gastroenterologist at the University Hospital, in São Paulo. [*Galileu* subcorpus]

12) Um gerente agrícola **AFIRMOU** que: 'Após a queima de imediato dá para entrar . . .' [Subcorpus Ciências Biológicas]

12) An agribusiness manager **STATED** that: 'After a burn it is possible to enter immediately . . .' [Biology subcorpus]

13) Os mais enérgicos **DIRÃO**: 'É fácil você se encantar quando fala nos bandidinhos . . .' [Subcorpus Ciências Humanas]

13) The more assertive **WILL SAY**: 'It is easy to be seduced when you talk about little thieves . . .' [Human Sciences subcorpus]

The projecting clauses, underlined in the examples below, were also classified according to their position in relation to the quoted discourse as occupying Theme position, like example 14, or Rheme, like example 15. Fewer than two percent of the occurrences of projecting clauses were found as occupying an intermediate position as in example 16 and were not considered so that the system could also be treated in binary terms:

14) Neste sentido, há entrevistados que DIZEM que a 'parte ruim são as instalações'. [Subcorpus Ciências Biológicas]

14) In this respect, there are interviewees who SAY that the 'negative aspect are the dependencies'. [Biology subcorpus]

15) ' . . . explosão das companhias tipo "ponto-com" tenha passado, há cada vez mais trabalho', **EXPLICA** Leonardo Leite, 37, coordenador da área de internet do escritório . . . [Subcorpus *Galileu*]

15) ' . . . the boom of the dot-com companies has passed, there is increasingly more work', **EXPLAINS** Leonardo Leite, 37, coordinator of the Internet area of the office . . . [*Galileu* subcorpus]

16) 'Muito bem', **DISSERAM** muitos dos colegas dele, 'então prove que estamos realmente . . .' [Subcorpus *Folha de São Paulo*]

16) 'Very well', **SAID** many of his colleagues, 'now prove we are really . . .' [*Folha de São Paulo* subcorpus]

The projecting clauses were finally classified according to the logical interdependency they establish with the quoted discourse. Although the most frequent logical interdependency relation between clauses in a complex of direct discourse representation is the paratactical one, as in example 17 below, given the occurrence of projection of direct discourse in hypotaxis, as in example 18, we decided to cover this phenomenon as well.

17) Em 1960, Mather (8) já afirmava: 'A determinação da bilirrubina é a mais irreal das dosagens bioquímicas' [Subcorpus Ciências da Saúde]

17) In 1960, Mather (8) already stated: 'The determination of bilirubin is the most unreal of chemical dosages'. [Health Sciences subcorpus]

18) Henrique Gandelman, em seu livro *De Gutenberg à Internet*, afirma que 'as perguntas se sucedem e as respostas nem sempre estão conseguindo . . .' [Subcorpus Ciências Sociais]

18) Henrique Gandelman, in his book *From Gutenberg to the Internet*, says that 'the questions are recurrent and the answers are not always able to . . .' [Social Sciences subcorpus]

By observing the systems of direct discourse representation chosen earlier we want to contribute to a better understanding of the ways scientific texts and texts about science construe interpersonal and social cultural meaning, negotiating discursive distance among different voices in discourse. In this article, we are specifically looking for analogies between and idiosyncrasies characteristic of each genre in the representation of direct discourse. We contend that, in the specific area of the discourse representation continuum we have chosen to examine, which seems to maximally accent discursive distance between the quoting and quoted voices, different genres will present different frequency profiles. The research article and the science popularization article are distinct genres interconnected by an intertextual potential. As such, they constitute different vehicles of discursive practice which make scientific knowledge circulate in different social spaces.

#### *4. Corpus and methodological procedures*

The study reported here is based on an electronic corpus of 34 million words compiled for the database of CORDIAL – Corpus for the Analysis of Language and Literature.<sup>1</sup> The texts analyzed in this article were downloaded from online sources and stored as text files prepared to be accessed by the software *Wordsmith Tools 4*. The electronic texts were checked and found to be, as far as text format files, analogous versions of the printed publications. The printed and online versions actually differ in terms of semiotic modes other than the verbal one, but these differences are not here discussed as they are beyond the scope of this article. The corpus is divided in two subcorpora, according to the genres investigated.

The research article corpus was compiled from SCIELO – *Scientific Electronic Library on Line* – website [[www.scielo.br](http://www.scielo.br)]. The 6490 items come from the Humanities, Social Sciences, Linguistics, Biology, Health Sciences, Engineering and Agrarian Sciences and total 34,400,383 words. The popularization article corpus was compiled from three Brazilian online sources: i) the science popularization magazine *Galileu* [<http://revistagalileu.globo.com>]; ii) the science section of the newspaper *Folha de São Paulo* [[www.folha.uol.com.br](http://www.folha.uol.com.br)]; iii) the Science and Technology section in the general interest magazine *Isto É* [[www.terra.com.br/istoe/](http://www.terra.com.br/istoe/)]. The 1174 items of the popularization article corpus amount to 893,991 words in total.

Our study made use of the software *Wordsmith Tools 4*, by Mike Scott [<http://www.lexically.net/wordsmith/>] to obtain quantitative information from the corpus. Annotation was done manually. Applying the utilities of the software, the initial corpus was searched so that we could compile a more restricted corpus, this time specifically including occurrences of direct discourse representation projected by reporting verbs and their co-text. The research article initial corpus provided 3012 occurrences of direct discourse representation with reporting verbs. The science popularization article corpus provided 2408 occurrences. Each of them was manually annotated according to: i) primary tense of the reporting verb:

present or past; and ii) thematic position of the projecting clause: Theme or Rheme; and iii) the logical interdependency relation established between the projecting clause and the quoted discourse: parataxis or hypotaxis. The occurrences were also classified according to the reporting verb categories suggested by Thompson and Ye (1991), but for the sake of space restrictions the analysis of reporting verbs will not be discussed here. For a complete discussion see de Oliveira (2005).

We then generated the systems' frequency profiles by interrogating the corpus in the following way: given X instances of reporting verbs in the corpus, which had either to be ascribed to the present or past tense, though not to both, how many occurrences of each tense could be found? And given 10,000 words of all lexical items present in the corpus, how many times on average can we expect to find an occurrence of a reporting verb in direct discourse in the present tense and in the past tense? This procedure was carried out for the three systems selected for analysis in the research article corpus and in the popularization article corpus. The frequency profiles of the two corpora were then contrasted. Due to space limitations, we will only briefly comment on the categories of the two most frequent reporting verbs in the corpus.

### 5. Data analysis

This section presents the frequency profiles of the two corpora under study regarding the three systems involved in direct discourse representation with reporting verbs: primary tense, Theme and logical interdependency. Table 1 presents information about the research corpus and its subcorpora. The subcorpora are ordered from the one which presents the lowest frequency of direct discourse representation with a reporting verb to the subcorpus which presents the highest frequency. Note that the last row of Table 1 summarizes the data of the initial research article corpus by presenting the sum of the subcorpora which constitute it.

TABLE 1. *Direct discourse representation with a reporting verb in the research corpus and its subcorpora*

<i>Raw corpora</i>	<i>Total number of words in the initial corpus</i>	<i>Occurrences of reporting verbs in direct discourse representation</i>	<i>Frequency of direct discourse representation in raw corpora per 10,000 words</i>
Agrarian sciences	2,957,367	3	–
Engineering	4,514,072	3	–
Health sciences	2,873,520	42	0.01
Linguistics	566,842	99	0.03
Biology	13,587,211	225	0.07
Social sciences	2,222,829	352	0.10
Human sciences	7,678,542	2288	0.67
Research article corpus	34,400,383	3012	0.88

TABLE 2. *Direct discourse representation with a reporting verb in the popularization article corpus and its subcorpora*

<i>Raw corpus</i>	<i>Total number of words in the initial corpus</i>	<i>Occurrences of reporting verbs in direct discourse representation</i>	<i>Frequency of direct discourse representation in raw corpora per 10,000 words</i>
<i>Galileu</i>	274,503	379	4.24
<i>Folha de São Paulo</i>	408,498	564	6.31
<i>Isto É</i>	210,983	1465	16.39
Popularization article corpus	893,984	2408	26.94

The results show clear disciplinary differences in the extent to which writers rely on direct discourse representation with a reporting verb. While writers in Agrarian Sciences and Engineering hardly employ this rhetorical resource, Health Sciences, Linguistics and Biology can be said to form a group of similar relative frequency, and Social and Human Sciences are the disciplines displaying the highest number of occurrences. The figures support Hyland's (1999) research, in the ordering of Engineering, Biology and Social Sciences regarding the frequency of citation.

Table 2 indicates the differences in the extent to which writers in the subcorpora of the popularization article corpus rely on direct discourse representation with a reporting verb. Note that the last row of Table 2 summarizes the data of the initial popularization article corpus by presenting the sum of the subcorpora which constitute it.

Note that *Galileu*, which among the popularization subcorpora is the only source entirely devoted to science popularization, is the subcorpus that presents the lowest frequency of direct discourse representation with reporting verbs. The newspaper *Folha de São Paulo*'s frequency profile is more similar to *Galileu*'s than to the general interest magazine *Isto É*'s, in which the frequency of the rhetorical resource here observed is twice the amount of the two other subcorpora together.

When we compare the frequency of occurrence of direct discourse representation with a reporting verb in the research article corpus and the popularization article corpus, figures clearly indicate this rhetoric resource plays a much more significant role in science popularization texts than in research articles. The research article corpus presents a frequency of occurrence of direct discourse representation with a reporting verb of 0.88 in 10,000 words, while the popularization article corpus presents an average of 26.94 occurrences per 10,000 words. These figures mean that direct discourse representation with a reporting verb is approximately 30 times more frequent in the popularization article corpus than in the research article corpus. In probabilistic terms, it takes approximately 100,000 words in the research article corpus to identify nine occurrences of reporting verbs projecting direct discourse, while, for the same

TABLE 3. *Probabilistic profiles of the systems of direct discourse representation with a reporting verb*

Corpus	Primary tense, present: past	Thematic position, Theme: Rheme	Logical
			interdependency, parataxis: hypotaxis
Research article corpus	0.70: 0.30	0.93: 0.07	0.72: 0.28
Popularization article corpus	0.80: 0.20	0.03: 0.97	0.99: 0.01

number of words in the popularization article corpus, we can identify approximately 269 occurrences of reporting verbs performing the same rhetorical function. The frequency profiles of the two corpora are clearly different regarding the occurrence of direct discourse representation with a reporting verb, the difference being higher than one order of magnitude.

We now turn to the analysis of the frequency profiles of the systems responsible for direct discourse representation with a reporting verb in the research and popularization corpora. Table 3 presents probabilistic figures of primary tense, thematic position and logical interdependency.

Within the system of primary tense, we notice both corpora favor direct discourse representation with a reporting verb in the present tense, the figures for the present tense being even higher in the popularization article corpus. Within the system of thematic position, figures in the two corpora mirror each other, that is, almost all projecting clauses of direct discourse in the research article corpus occupy thematic position, while in the popularization article corpus it is the rhematic position which becomes a highly favored option for the projecting clause. Turning to the logical interdependency relation established between the projecting clause and the projected discourse, we notice that, while in the popularization article corpus nearly all direct discourse is projected in parataxis, in the research article corpus approximately 30 percent of the projecting clauses enter in a logical relation of hypotaxis with the projected discourse.

The data about grammatical systems presented above indicate, on the one hand, that the frequency of occurrence of direct discourse representation with a reporting verb is a rhetorical resource that plays a much more significant role in the popularization article corpus than in the research article corpus and, on the other hand, that the systems involved in this phenomenon tend to asymmetry or extreme asymmetry in the popularization corpus, while allowing for more variation, redundancy in probabilistic terms, in the research article corpus. At the other end of the *lexicogrammatical* continuum, we can look at the reporting verbs observed in this research as lexical items, options in an open-ended system.

We classified the 5420 occurrences of reporting verbs projecting direct discourse identified in the corpora investigated according to the categories

suggested by Thompson and Ye (1991). Due to space limitations, we will just briefly discuss some of the most relevant results concerning the following data: i) 206 different verbs were identified functioning as reporting verbs; ii) among the 5420 occurrences of reporting verbs in direct discourse representation and the 206 different possibilities of lexical options, 2007 occurrences accounted for the choice of one of the grammatical forms of the verb '*afirma*' ('states') and 972 occurrences for the choice of the verb '*diz*' ('says'). Viewed from the perspective of Thompson and Ye's model of analysis, approximately 55 percent of all occurrences of the reporting verbs in direct discourse representation indicate, through the two most frequent lexical realizations found in the corpus – *afirmar* and *dizer* – the writers' necessity to transfer to the authors the responsibility for the quoted information, choosing to represent it as a textual/verbal act, and, within their evaluation potential, the writer's difficulty to expose an explicit evaluation of the quoted information, choosing to present it more objectively with so-called more neutral reporting verbs, such as '*say*'.

## 6. Discussion

We take the research article and the popularization article corpora here investigated as representative of the genres research article and science popularization article. We contend that the frequency of direct discourse representation with reporting verbs in both genres can be interpreted according to the same principle: the more accented the discursive asymmetry between the quoting and quoted voices in discourse, the higher the frequency of direct discourse representation with reporting verbs. In other words, besides the substantially discussed functions attributed to direct discourse representation, of conferring texts more credibility and dramatization, this research suggests direct discourse representation with reporting verbs should be interpreted in at least two dimensions. First, it signals through the experiential component of the ideational metafunction the rhetorical pressure writers feel to accent the presence of different voices in discourse, signaling asymmetrical authority relations between the quoted and quoting voices. Second, direct discourse representation, within the interpersonal metafunction, places writers and readers in interactive relation, opening up evaluative space which distances the quoted and the quoting voices and didactically orientating readers' interpretation toward writers' preferred reading of the quoted discourse. Note, however, that the existence of evaluative space does not imply evaluation will be readily carried out by writers, which casts doubt on the observation of reporting verbs alone as a key rhetorical resource for investigating evaluation in texts.

Our own perception from the Brazilian corpus we have investigated is that, between the poles of discourse representation, direct and indirect discourse, the continuum of representation negotiates discursive distance between writers and the authors they quote. Such distance signals asymmetrical authority relations in discourse. Thus, direct discourse accents distance, structurally marking the

quoted voice as alien, different from the text writer's voice. Indirect discourse minimizes distance, as writers feel confident enough to take possession of the original discourse and present it under their own guise. The continuum of discourse representation could then be thought of as composed of different and innumerable layers of fusion between the voices of writers and authors.

Obviously enough, since the decision to use direct or indirect discourse, or any of the layers of the continuum between the two, is the text writer's, the effect of fusion between the quoted and the quoting voices is rhetorical. There is no real fusion, but the presentation of different voices in a more symmetrical or asymmetrical discursive relation. What we want to emphasize is that the more asymmetrical the authority relation between the voices in a text, the greater the need to accent discursive distance and signal explicitly whose voice belongs to whom. Conversely, the more symmetrical the authority relationship between the writer and the quoted authors, the more comfortable the writer will feel to present the quoted voice as dependent on her own, in a structural and semantic way.

It is possible to claim that both the research article and the science popularization article place the voices of the writers as in an inferior authority position as regards the voices they quote. This claim relies on the facts that i) in both genres the reporting verbs selected to accent discursive distance signal the need to transfer to the author the responsibility for the quoted information and the difficulty writers feel to explicitly express an evaluation of the quoted discourse; ii) through the frequent selection of the present tense for the reporting verbs writers seem to remove the represented discourses from a time perspective, conceding them the *status* of facts and indisputable truths; and, finally, iii) the most frequent logical interdependency established between the quoting and the quoted discourse is parataxis, that is, the logical semantic relationship that most clearly signals the separation of the voices of writers and authors is independent clauses, thus accenting discursive distance between quoted and quoting voices.

Now, if all the aspects mentioned above show that the genre's research article and the science popularization article resemble each other, note that in the former the discursive distance between the voices is less accented than in the latter. In the research article corpus, almost 30 percent of direct discourse representation with a reporting verb was projected through hypotaxis. That is, the research article not only selects the layer of the continuum of discourse representation that more clearly indicates separation between voices in discourse 30 times less frequently than the science popularization article, but it also rearranges the grammar of direct discourse projection so as to minimize discursive distance between writers and the quoted authors. Such a manipulation of resources within the ideational metafunction instantiates interpersonal meanings specific of the academic circles, where one of the most relevant causes of asymmetry among writers, authors and readers could be said to be the experience in the research field.

The observation of the thematic system involved in the projection of direct discourse representation with a reporting verb as shown in the data analysis

section seems to confirm this interpretation. We consider a projecting clause in a thematic position emphasizes the writer's representation of the quoted discourse, while a projecting clause in a rhematic position emphasizes the *status* of the quoted voice and the quoted discourse itself. In other words, a projecting clause in thematic position can be understood to put emphasis on *how* the quoted discourse fits into the discourse of the writer. Alternatively, a projecting clause in a rhematic position can be thought to put emphasis on *what* the quoted voice has to say and on *who* actually says it. In the research article corpus 93 percent of the projecting clauses are thematized, while in the popularization article corpus 97 percent of the projecting clauses are in rhematic position. As a result, if in the research article asymmetrical differences between the voices in discourse are marked by the selection of reporting verbs which transfer to the quoted voice the responsibility for the information represented and not explicitly express evaluation of it, in the science popularization article such differences are accented by, besides the selection of lexical realizations of reporting verbs broadly similar to the research article, a grammatical structuring that increases discursive distance.

## 7. Conclusion

We have suggested that the projection of direct discourse by a reporting verb can be interpreted as simultaneously signaling asymmetrical authority relations in discourse and serving as a rhetorical didactical resource that writers may use to favor certain readings of the discourses they quote. We have interpreted the approximately 30 times more frequent occurrence of direct discourse representation with reporting verbs in science popularization articles than in research articles as an indication that the voices quoted by journalists in texts about science are placed in a discursive position superior to their own.<sup>2</sup> Likewise, the frequency of reporting verbs projecting direct discourse in research articles is understood as signaling asymmetrical, but less accented discursive distance relations between writers and the authors they quote.

If we accept the claim that in the research article the discursive distance between writers and the authors they quote is less accented than in the science popularization article, we can expect writers in the former to have more discursive freedom to appropriate the discourses they quote. Such a freedom is intrinsically related to the intended audiences of each genre. While writers in the academic circles write, primarily to their peers, journalists, on the other hand, write to other journalists, lay people, scientists, and scientists involved in popularization processes. This could possibly explain why the layer of the continuum of discourse representation that most clearly separates the quoted and quoting voices in discourse is 30 times more frequent in the science popularization article than in the research article. Likewise, it would explain why, among the popularization sources investigated, the one magazine which

specializes in science popularization, *Galileu*, was the subcorpus which presented the lowest frequency of direct discourse representation with reporting verbs.

The use of direct discourse representation with a reporting verb as a key rhetorical resource in the science popularization article can be understood in at least three complementary dimensions. First, it construes the quoted authors, that is, those who produce scientific knowledge, as superior to writers, those who popularize it. Second, it limits journalists' possibilities of appropriating the voices they quote in their discourse. Third, though discursive distance opens up the possibility of evaluation, it does not contribute to the subversion of social and cultural differences, since the expression of explicit evaluation is limited by the genre's rhetorical conventions. It is suggested that the *status* culturally and historically ascribed to scientific knowledge and scientists prevents journalists from occupying the evaluative space opened up by discursive distance more comfortably, while making pressure for the clear identification of which voice belongs to whom in discourse.


## NOTES

1. CORDIALL – Corpus of Discourse for the Analysis of Language and Literature – is the name of the corpus that began to be compiled in 1999 by a group of researchers at the Faculdade de Letras of the Federal University of Minas Gerais, Brazil. For more information on the corpus compilation, see Alves et al. (2004).
2. We would like to acknowledge and thank the referees of this article, who made detailed and useful criticisms as well as contributions, such as pointing out that the high frequency of direct speech with reporting verbs in science popularization articles may be correlated to the fact that these articles draw more frequently on interviews rather than written texts. Direct quotes are traditionally associated to the signaling of more immediate and vivid interaction.

## REFERENCES

- Alves, E., Pagan, A. and Magalhães, C. (2004) 'Towards the Construction of a Multilingual, Multifunctional Corpus: Factors in the Design and Application of CORDIALL', *TradTerm: Revista do Centro Interdepartamental de tradução e terminologia* 10: 163–78.
- Biber, D., Johansson, S., Leech, G., Conrad, S. and Finegan, E. (1999) *Longman Grammar of Spoken and Written English*. London: Longman.
- Caldas-Coulthard, C.R. (1994) 'On Reporting Reporting: The Representation of Speech in Factual and Factional Narratives', in M. Coulthard (ed.) *Advances in Written Text Analysis*, pp. 295–320. London: Routledge.
- Calsamiglia, H. and Ferrero, C.L. (2001) 'Polifonía en textos periodísticos con información científica', *Lengua, discurso, texto: I Simposio internacional de análisis del discurso* V2: 2647–63.
- Calsamiglia, H. and Ferrero, C.L. (2003) 'Role and Position of Scientific Voices: Reported Speech in the Media', *Discourse Studies* 5(2): 147–73.
- Calsamiglia, H., Bollina, H., Cassany, D., López, C. and Martí, J. (2001) 'Análisis discursivo de la divulgación científica', *Lengua, discurso, texto: I Simposio internacional de análisis del discurso* V2: 2639–46.

- Crismore, A. and Farnsworth, R. (1990) 'Metadiscourse in Popular and Professional Science Discourse', in W. Nash (ed.) *The Writing Scholar: Studies in Academic Discourse*, pp. 118–36. London/Newbury Park, CA/New Delhi: Sage.
- Fairclough, N. (1992) *Discourse and Social Change*. Cambridge: Polity Press.
- Halliday, M.A.K. (1998) 'Things and Relations: Regrammaticising Experience as Technical Knowledge', in J.R. Martin and R. Veel (eds) *Reading Science: Critical and Functional Perspectives on Discourses of Science*, pp. 185–235. London/New York: Routledge.
- Halliday, M.A.K. (2004) *The Language of Science*. London/New York: Continuum.
- Halliday, M.A.K. (2005) *Computational and Quantitative Studies*. London/New York: Continuum.
- Halliday, M.A.K. and Martin, J.R. (1993) *Writing Science: Literacy and Discursive Power*. London/Washington, DC: University of Pittsburgh Press.
- Halliday, M.A.K. and Matthiessen, C.M.I.M. (2004) *An Introduction to Functional Grammar*. London: Arnold.
- Hyland, K. (1999) 'Construction of Disciplinary Knowledge', *Applied Linguistics* 20(3): 341–67.
- Leech, G.N. and Short, M.H. (1981) *Style in Fiction*. London: Longman.
- Myers, G. (1990) *Writing biology: texts in the social construction of scientific knowledge*. London: University of Wisconsin Press.
- Myers, G. (2003) 'Discourse Studies of Scientific Popularization: Questioning the Boundaries', *Discourse Studies* 5(2): 265–79.
- de Oliveira, J.M. (2005) *As vozes da ciência: a representação do discurso nos gêneros artigo acadêmico e artigo de divulgação científica* Unpublished PhD thesis: Universidade Federal de Minas Gerais, BR.
- Orlandi, E.P. (2001) *Discurso e Texto: formulação e circulação dos sentidos*. São Paulo: Pontes.
- Semino, E. and Short, M. (2004) *Corpus Stylistics: Speech, Writing and Thought Representation in a Corpus of English Writing*. London: Routledge.
- Swales, J.M. (1985) 'Citation Analysis and Discourse Analysis', *Applied Linguistics* 7(1): 39–56.
- Swales, J.M. (1990) *Genre Analysis: English in Academic and Research Settings*. Cambridge: Cambridge University Press.
- Swales, J.M. (2004) *Research Genres: Exploration and Applications*. Cambridge: Cambridge University Press.
- Thompson, G. (1996a) *Helping Learners with Real English*. London: HarperCollins.
- Thompson, G. (1996b) 'Voices in the Text: Discourse Perspectives on Language Reports', *Applied Linguistics* 17(4): 501–30.
- Thompson, G. and Ye, Y. (1991) 'Evaluation in the Reporting Verbs Used in Academic Papers', *Applied Linguistics* 12(4): 365–82.
- Zamboni, L.M.S. (2001) *Cientistas, jornalistas e divulgação científica: subjetividade e heterogeneidade no discurso da divulgação científica*. São Paulo: Autores Associados.
- White, P.R.R. (2003) 'Beyond Modality and Hedging: A Dialogic View of the Language of Intersubjective Stance', *Text* 23(2): 259–84.

 JANAINA MINELLI DE OLIVEIRA holds a PhD in Applied Linguistics from Universidade Federal de Minas Gerais, Brazil, and is presently working as a researcher at the

Observatorio de la Comunicación Científica, UPF, Spain. She has been a visiting researcher at Universitat Autònoma de Barcelona with a scholarship from the Brazilian agency Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - CAPES. Her current research interests are systemic functional linguistics, probabilistic approaches of grammar, scientific writing and popularization. Her article 'Science Update in Brazilian Television News: Social Actions of a Cultural Product' will be published in the Proceedings of the First International Conference on Critical Discourse Analysis held in Valencia, Spain, in 2004. ADDRESS: c/ Rocafort 129, 1<sup>o</sup>-3<sup>a</sup> Barcelona, Catalunya 08015, Spain. [email: janaina-minelli.deoliviera@upf.edu]

ADRIANA PAGANO is Professor of English and Translation Theory at Universidade Federal de Minas Gerais, Brazil. She holds a PhD in Comparative Literature from Universidade Federal de Minas Gerais and has been a visiting scholar at the University of Nottingham, UK and the University of Massachusetts at Amherst, USA. Her current research interests are systemic functional linguistics, genre analysis and translation studies. She has co-authored *Traduzir com Autonomia: estratégias para o tradutor em formação* (Contexto, 2000), *Competência em tradução: cognição e discurso* (Editora da UFMG, 2005) and edited *Metodologias de pesquisa em tradução* (FALE, 2001). [email: pagano@netuno.lcc.ufmg.br]