Dialogical surface text features in abstracts

Ingrid García-Østbye
E.O.I. Quart de Poblet, Valencia (Spain)
ingarost@postal.uv.es

Abstract

A sample driven description of Research Article-Comment-Reply (RA-C-R) abstracts in terms of abstract sentence length, reference, possessive structures, modal verbs and word range was carried out to find out whether their surface text features showed some trace of a dialogical construction of knowledge within the psychology discourse community. The study served an exploratory purpose. A Boolean search was conducted in the PsycLIT database yielding a sample of 149 PsycLIT RA-C-R abstracts (13,978 words). Relative frequency percent distributions were calculated for all variables, including reported speech verbs. Specific comparisons with a Medline corpus were conducted and variations were accounted for in terms of scientific discourse characteristics, field, database policies, and dialogical nature; that is, in the framework provided by the strands of research of quantitative applied linguistics, social concerns in genre analysis and the model monopoly theory developed in the implementation in sociology of the systems theory. The results suggest: (i) a word range affected by both psychology as a discipline and the dialogical content on which PsycLIT RA-C-R abstracts report; (ii) a complementarity of reference and possessive structures characterised by features of scientific discourse, feedback genres and dialogical dimensions; (iii) the presence of both deontic and epistemic modality in the modal verbs of our sample; (iv) and also that abstract length, sentence length and number of sentences per paragraph in our sample may not vary greatly in general terms from those of the social sciences.

Keywords: corpus linguistics, text features, dialogue, reply abstract, reported speech.

Resumen

Características textuales superficiales dialógicas en resúmenes tipo "abstract"

Procedimos a realizar una descripción de una muestra de "resúmenes tipo abstract de las respuestas a los comentarios sobre los artículos" (RA-C-R) de la base de datos PsycLIT en términos de longitud de la oración, referencia, posesivos, modales y abanico terminológico, con el objetivo de averiguar si las características textuales superficiales de dichos resúmenes revelaban huella de la construcción dialógica de conocimiento en el seno de la comunidad de discurso de psicología. Se trata de un estudio de carácter exploratorio. Se realizó una búsqueda booleana que rindió una muestra de 149 resúmenes RA-C-R (13.978 palabras). Se calculó las distribuciones de frecuencia relativa de todas las variables, incluyendo verbos de estilo indirecto y se realizaron comparaciones especificas con un corpus de Medline explicando las variaciones encontradas en términos de característica del discurso científico, disciplina, norma de base de datos y naturaleza dialógica; es decir, en el marco proporcionado por las tradiciones de la lingüística cuantitativa aplicada, las consideraciones sociales en el análisis de género y la teoría de monopolio de modelo de la teoría de sistemas en sociología. Los resultados indican que los resúmenes muestran: (1) un abanico terminológico influido tanto por la psicología como disciplina como por el contenido de naturaleza dialógica del que informan los resúmenes; (2) los resultados revelan la presencia de una complementariedad de las estructuras de referencia y posesivas, caracterizada por rasgos del discurso científico, géneros de feedback y aspectos dialógicas; (3) la presencia de modalidad deóntica y epistémica en la muestra y (4) la longitud del resumen, la longitud de la oración en el resumen y el número de oraciones por párrafo de la muestra no difieren en gran medida en términos generales de las encontradas en las ciencias sociales.

Palabras clave: lingüística de corpus, características textuales, diálogo, abstract de la respuesta, estilo indirecto.

1. Introduction

The strands of research of genre analysis and systems theory in sociology might be said to address and share the common ground of the scientific construction of knowledge. This may be seen as both an individual and social task, an on-going, never-ending dialectical process taking place generation after generation. In this process, the evaluation of new knowledge and its acceptance, and the communication of knowledge itself, have adopted both spoken and written discourse forms, such as

conversations, dialogues, comments, remarks, letters, drafts, chats, research articles, referee comments and replies, among others. Conferences, seminars, universities, journal chat rooms as well as abstracting services, databases and journals provide the milieu for the myriad events that allow for the construction of knowledge within scientific communities; among these PsycLIT replies to comments on articles might be included.

The flowchart in Figure 1 illustrates the process of writing a research article (RA), a comment and a reply, together with their abstracts, which takes place within discourse communities.

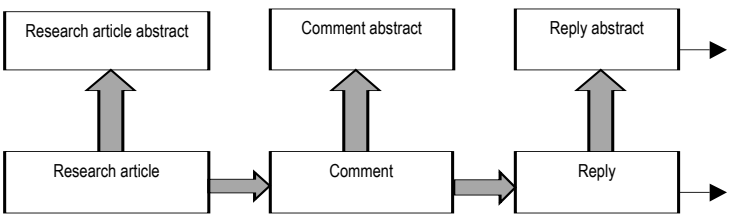

Figure 1. The process of writing and abstracting RAs comments and replies.

In this process "fully invested disciplinary actors" (Berkenkotter & Huckin, 1995: 25) or "model strong actors" (Bråten, 1981: 160) seem to play a major role as authors, referees, reviewers, gatekeepers and expert judges. Genre analysis has conceptualised discourse communities as socio-rhetorical networks whose established members possess familiarity with the particular genres (Swales, 1990). It has also paid attention to fully invested actors' awareness of this process of accommodation of new knowledge and its effect on the community genres and participation demands (Berkenkotter & Huckin, 1995), as the mastery of conventions seems to be a major condition for participation and eligibility, and has studied how authors, editors and reviewers transform manuscripts into validated knowledge claims (Chubin & Hackett, 1990). Peer review has been considered to have a range of functions including experts' social control of the quality of new contributions entering the field; this confers on them authority and authenticity (Berkenkotter & Huckin, 1995), and they can be considered an example of status judges charged with evaluating the quality of role performance in a social system (Merton & Zuckerman, 1973).

The shaping of both discourse and knowledge, gatekeeping and eligibility for publication has been addressed referring to variables such as

social/epistemological discipline shifts (Bazerman, 1988), composition strategies (Myers, 1990), contextual relevance and register awareness (Sionis, 1995) and the relationship between discourse and the cognitive model for the discipline (Love, 1993). New developments in computer sciences have also introduced new approaches to the study of text and they have gone far beyond what Huddleston (1971) had envisioned when he published his book on *The Sentence in Written English* (see Swales, 2004). In recent years many corpus based studies have been published, particularly on RAs as well as abstracts (Hyland, 2000; Ferguson, 2001; Upton & Connor, 2001; Liebscher & Groppe, 2003; Swales, 2004; Piqué-Angordans & Posteguillo, 2006; among others).

The presence of dialogical dimensions and feedback has also raised concern in works regarding the distinction between informative and critical abstracts (Lancaster, 1991), book reviews (Motta-Roth, 1998), letters to the editor (Berkenkotter & Huckin, 1995), and a dialectical dimension of genres (van Peer, 1990); efforts which emphasize the social knowledge shared by participants, the dialogical nature of the genre and the dialectical interpenetration of subjective and objective aspects.

Bråten (1981) addresses the diachronical and synchronical epistemological cognitive functioning of communities in general, and scientific communities more specifically, as regards to cognitive interaction:

> […] scientists […] carry out operations involving symbols and representations within a self-enclosed domain which they maintain qua living systems. Their understanding is achieved through their involvement in a more or less shared culture of symbols and artefacts. This shared understanding is intersubjective –neither subjective, nor objective– and it generates the subject-object complementarity. (Bråten, 1981: 160)

In this sense Bråten (1981) analizes if (relevant) members of the academy always direct and help the scientific community properly reach their (adequate) goals. He considers this question from the point of view of three system positions (as shown in Figure 2), which approach cognitive interaction respectively in (A) "Artificial terms of symbolic representations", (B) "Biological terms of autopoiesis" –or more generally organisational closure, and (D) "Dialogical terms of discourse and dualities", including the pair of modes conforming to both (A) and (B).

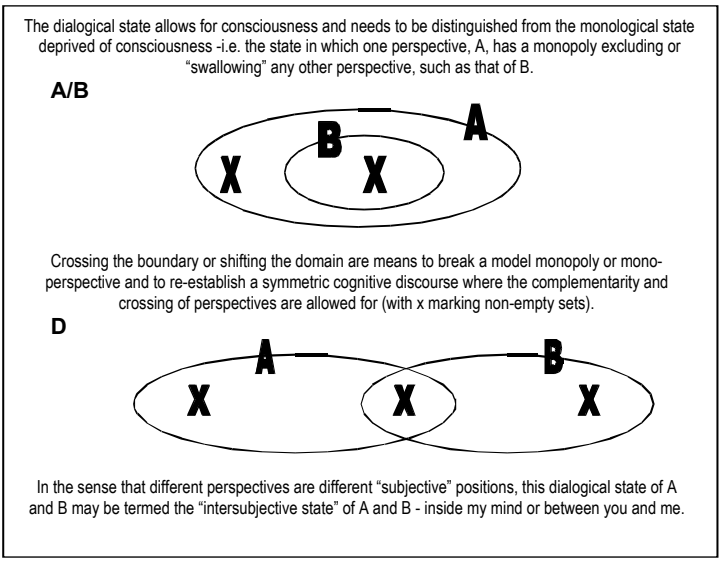

Figure 2. The model power thesis of interaction (Bråten, 1981: 160)

In (A) approach, a sociocultural system is defined as a meaning-processing system of interacting participants who maintain and transform the identity of themselves and that of their network through a more or less shared understanding of both themselves and the world. Under certain conditions, this shared understanding or world view may become monolithical and closed to such a degree that it rules out any rival view, and thereby prevents dialogue in a symmetric sense. This systems state may be called a "model monopoly" in terms of the first position (A), or a "consensual mono-perspective" in terms of the second position (B). In his perspective, Bråten (1981) considers that a given universe of discourse, E, may be predefined in such a manner that only one of the two participant actors, A, is rich in relevant concepts and symbolic representations which reflect his own interests and perspectives. Therefore, A is the actor known as "the model-strong actor":

> If all the elements and relations in E which are describable in terms of B's perspective, also are describable in terms of A's perspective, and there are elements in E that are describable only in A's but not in B's perspective, then A is the model-strong actor, and B *the model-weak one*, with respect to E. (Bråten, 1981: 162)

The perspective of B can be "swallowed" by –or properly included– in A's perspective; in our context, in Swales' (1990) terms, this is frequent in the case of an expert holding a conversation with a student or a "novice" member of the discourse community. The conditions for the resolution of the model monopoly are the concern of the third position (D) proposed by Bråten; the Dialogical Gestalt of conversation and complementarities.

> […] this dialogical state of A and B may be termed the 'intersubjective state' of A and B –inside my mind or between you and me. There is organizational closure and conceptual agreement in the intersection, but with requisite inconsistency of perspectives. (Bråten, 1981: 162)

Following social concerns in genre analysis as well as Bråten's (1981) and Swales' (1990) perspectives, the corpus of PsycLIT RA-C-R abstracts used in this study and the literature they report on might be said to:

(i) be likely to be studied to eventually be taught on how to provide feedback in the psychology discourse community;

(ii) be used by accomplished/novice members of the psychology discourse community in their participation; being a final product of a process of gatekeeping and eligibility; hence, being shaped to meet the demands and expectations of future accomplished/novice readers, editors and referees;

(iii) be one of the (research-related) subgenres the psychology community possesses, showing the lexis employed by the psychology discourse community; and also a written expression of the psychology discourse community cognitive functioning, in the overlapping of written genres and dialogical forms;

(iv) be present in journals and databases, as mechanisms of intercommunication among the members of the psychology discourse community in the awareness and provision of feedback; hence, they may provide an arena for discussion of epistemological concerns in psychology in their construction of knowledge, express the presence of evaluation, acceptance and accommodation of new knowledge, assess the negotiation of knowledge claims, help the fitting of research into the established field, and contribute to the validity, prevalence and resolution of models in time.

2. Research questions and aim of this study

An initial description of PsycLIT RA-C-R abstracts is offered in order to answer the following research questions: Are PsycLIT RA-C-R abstracts different from other abstracts? And if so, can their differences be accounted for in terms of a dialogical nature, and to what extent?

These questions are considered in the current study, whose main objective is to contribute to an initial description of PsycLIT RA-C-R abstracts as regards to basic features such as abstract length, average sentence length, pronoun usage, possessive structures, modal verbs and reported speech verbs and word types; a preliminary study paying attention to those surface linguistic features that might suggest the presence of a dialogical construction of knowledge within the psychology discourse community, in lexis such as nouns and reporting structures, and modality –whose mastery seems important in eligibility for publication.

In other words, I aim to disclose whether the sample of PsycLIT RA-C-R abstracts differs from other abstracts and scientific discourse in general in a set of relevant features, and analize whether this might be attributable to a dialogical construction of knowledge within the psychology discourse community. More specifically the research questions were:

a) Are PsycLIT RA-C-R abstracts and their average sentence the same length as other abstracts?

b) What is the range and percent distribution of the personal pronouns used?

c) What is the range and percent distribution of the possessive structures used?

d) What is the range and percent distribution of modal verbs?

e) What is the range and percent distribution of the reported speech verbs used?

f) Do PsycLIT RA-C-R abstracts show any other differences in lexis as regards to range or percent distribution?

A sample PsycLIT RA-C-R abstract is reproduced in the appendix for further reference and discussion purposes.

3. Method of Study

A sample of PsycLIT RA-C-R abstracts was observed and specific comparisons with a Medline corpus were conducted. A multiple perspective was adopted by analyzing texts as regards to different variable features.

i) number of words;

ii) number of sentences;

iii) range and number of personal pronouns;

iv) range and number of possessive structures;

v) presence or absence of subject;

vi) range and number of word types;

vii) range and number of modal verbs;

viii) range and number of reported speech verbs.

Reference was considered a suitable starting point as the observation of the sample of PsycLIT RA-C-R abstracts, particularly of their initial sentences suggested that the Saxon genitive, other possessive structures and subjectless sentences might be acting complementarily. The whole range of word types was registered and some specific words –including reported speech verbs– were also observed on the grounds that PsycLIT RA-C-R abstracts –as feedback abstracts– might show some trace of a dialogical construction of knowledge in their surface text features. Relative frequencies were also calculated for each word and factor in each corpus and for the range of items sampled within each variable over its total figure of occurrences –yielding an inner percentage distribution. Specific comparisons with the Medline corpus were then conducted.

The main criteria for the selection of the sample abstracts in the PsycLIT corpus were their representativity, as they were all included, and the reputation of the database.[1] The abstracts in our sample were selected by conducting a Boolean search in the PsycLIT database on CD ROM using the operators "research article" (RA), "comment" (C) and "reply" (R) to narrow the search down in the span 1996-98 first term.[2] The initial search on the provision of feedback yielded 5,978 RAs; 3,939 Cs; and 1,907 Rs. The corpus was made up of 149 PsycLIT RA-C-R abstracts matching the aforementioned search requirements. The corpus consisted of 13,978 occurrences and 1,104 sentences. The PsycLIT RA-C-R abstracts of our

sample constituted an analytic sample –i.e. it served an exploratory purpose.[3] The programme used to study the corpus was *WordSmith Tools* (Scott, 1996).

4. Results

The corpus of PsycLIT RA-C-R abstracts used a range of 3,106 different words out of a total of 13,978 occurrences. In our corpus the highest occurrence frequencies were shown by predeterminers such as "the" (f=966), "that" (f=307), prepositions such as "of" (f=586), conjunctions such as "and" (f=504); followed by auxiliary verbs "be" (f=382), "do" (f=458), "have" (f=49), representing 3% of the whole sample, and nouns such as "author" (f=78), "article" (f=99), "comments" (f=123), "patients" (f=38), "treatment" (f=22), "results" (f=26), "data" (f=25) and initials standing for first names such as J. (f=71), pronouns such as "they" (f=29), adjectives such as "original" (f=32), "clinical" (f=15) and adverbs such as "rather" (f=17). Our sample yielded 2,190 verb occurrences, representing a 15.66% of the whole corpus; auxiliary verbs (f=458) accounted for 3.2% of the occurrences, whereas reporting verbs (f=1587) accounted for an 11.35% of the whole sample and modal verbs (f=92) for a 6.5% of the total sample.

In the list of most commonly used words from our corpus words such as "comment", "reply", "respond", "study", and others, irrespective of their grammar class, are used quite frequently. For example, "comment" is the tenth runner-up word in terms of frequency, appearing 172 times[4], which accounts for 0.88% of use in reference to the total number of words of the corpus (13,978). In other words, only grammatical words, such as "the", "of", "and", "a/an", and the like, appear in the first positions. Table 1 gives a clear account of the first ten words that appear most frequently in the sample.

		f	%
1	the	966	6.91
2	of	586	4.19
3	and	504	3.61
4	to	384	2.75
5	a/an	354	2.53
6	that	307	2.20
7	in	290	2.07
8	on	215	1.54
9	by	178	1.27
10	comment, -s, -ed	172	1.23

Table 1. First ten most frequently used words in the corpus.

Further down the list, we can see other terms, that directly or indirectly refer to reporting verbs, such as "reply", with its different forms, appearing in the 28th position (f=57; 0.41%), "respond" in the 30th (f=52; 0.37%), etc. This shows a distinctive use of such terms in this type of abstracts.

4.1. Abstract and average sentence length

The RA-C-R abstracts of our corpus showed an average length of 107.9 words. The abstracts of our sample were slightly shorter than other abstracts reported in previous literature. This might be explained as a database policy, as degree of importance attached to the item summarized (feedback subgenres). Medline abstracts in our corpus were slightly longer (132 words per abstract); their summary contents might be different in nature; Medline abstracts reproduced purpose, methods, results and conclusions informatively, whereas PsycLIT RA-C-R abstracts reproduced the most salient arguments or the outline of the discussion referred indicatively and informatively.

In our sample the average number of sentences per paragraph was 3.41. This means that PsycLIT RA-C-R abstracts used fewer sentences than other abstracts, compared against the historical framework provided by diachronic studies of this feature in abstracts showing means ranging from 4.4 (in 1944) to 6.4 (in 1989). Medline abstracts showed an average 9.1 sentences per abstract.

The average sentence length was 27.5 words per sentence (SD=12.90). These findings indicate that the sentence length of our corpus is slightly higher than reported in previous findings –around 25 as reported by Swales (1990, quoting from Bazerman, 1988)–, but does not differ substantially from average reported mean scores across disciplines –displaying means ranging from 23 up to 24.9 (McDonald, 1990)– placing our sample above the average in this feature. In this sense, McDonald (1990: 35) states that "academic writing is characterized by longer sentences than […] fictional or journalistic writing and […] academic writers write sentences averaging 23-26 words per sentence". Piqué-Angordans & Coperías (1999) find average sentence lengths ranging from 19.68 to 27.06 (SD=12.95) when studying sentence length in health science RAs; this span, however, widens from 18.07 to 32.04 (SD=11.95) when studying scientific articles in general (Piqué-Angordans & Andreu-Besó, 2000).

4.2. Subjectless sentences

In general terms, subjectless sentences (f=113) represented 10.2% of the whole sample, whereas sentences with a subject (f=991) accounted for 89.7% of the sentences. This is shown in Figure 3.

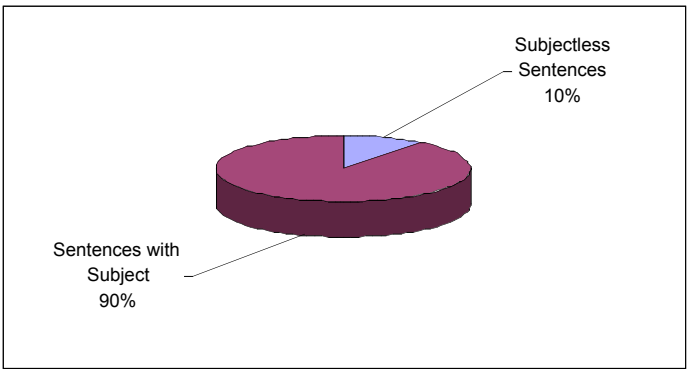

Figure 3. PsycLIT RA-C-R abstracts: subjectless sentences vs. sentences with a subject.

From the point of view of the abstracts themselves, 126 of them contained one or more subjectless sentences. This meant that the percentage of abstracts not including a subjectless sentence only reached 15%, and suggests the presence in our sample of the text of the AU (author) field of the abstract (AB) functioning as a "referent" subject, a typical feature of some abstracting journals, a matter of style. The PsycLIT RA-C-R abstracts initial subjectless sentence seems to constitute a first step in a progressive chain of reference, ranging from more specific to more general. For example:

Field	text
AU =	Smithson…
AB =	{subjectless} comments on […]
	the author […]
	s/he states that […]

The word "author/s" appears 130 times in the PsycLIT corpus (0.93%), while only 4 times in the Medline corpus (0.004%). In short, Medline abstracts do not show this reference pattern.

4.3. Personal pronouns, possessive adjectives and Saxon genitive

Personal pronouns were scarcely used in our corpus (f=111; 0.7%) and they were twice less frequently employed in the Medline corpus (f=548; 0.243%).

This can be explained as a preference for names and nouns on the part of the writer and suggests specificity and specialisation in both corpora. Reference to authors of RAs, Cs and Rs in the PsycLIT corpus, not present in Medline, may account for the difference found between them.

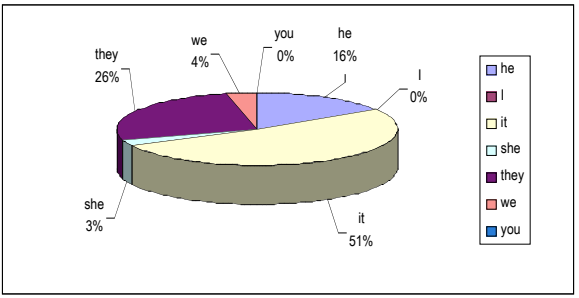

Figure 4. PsycLIT RA-C-R abstracts: personal pronouns.

As shown in figure 4, the absence of the first person singular pronoun "I", the second person singular pronoun "you" and the second person plural pronoun "you" was also observed in both corpora. This suggests the presence of impersonal style and attention to the authorship of abstracts. The third person reference in the personal pronoun distribution showed the highest incidence values; singular pronouns such as "he" (16.36%) or "she" (2.7%) were relatively rare, whereas "it" accounted for the highest percentage of pronoun occurrences (51.3%), as shown in the following excerpts.

The sample text [1] shows the use the personal pronoun "he", the name of the author of the reply ("Herzog") has been specified and "he" seems to avoid its repetition:

> [1] *Herzog* states that the amount of coverage in print media represents a fairly crude reflection of the cultural status of a grassroots social movement. However, *he* offers the meager turnout for the 1996 ... [*italics added*]

In general terms, these results could be explained as an avoidance of the "I-author/abstractor" and "you-reader" reference on the part of the writer of the abstract together with the presence of high levels of the pronoun "it" showing a range of functions ("It" passive, "empty *it*" subjects, including "cleft" sentences) in the PsycLIT sample as illustrated in excerpt [2]:

> [2] Replies to the B. P. Ryan and B. V. Ryan (1996) comment on [...]. *It* is argued that the notion that successful treatment rests on the foundation

> of operant conditioning is too simplistic, naive and outdated ... [*italics added*]

In example [2] the "*it*-passive", having no referent, might suggest a rhetorical choice, avoiding the authorship of the argument in this context, and implying a preference for impersonality. In our sample, "it" functions as a subject in "*it*-passives" and "cleft" sentences –presumably having a different rhetorical function– and as a way of avoiding the repetition of names and nouns. The example below shows how "it" may be used in our corpus, referring to (very often abstract) nouns and/or noun phrases and clusters and avoiding their repetition:

> [3] Replies to the M. D. Zeiler (1996) comment on [...]. Zeiler appreciates "the author's organism-centered approach" but believes that *it* issues from the wrong framework and focuses on the wrong phenomena. [*italics added*]

The third person plural pronoun, "they" (26%), was the second pronoun most frequently used in our corpus. The following text is an example of the use of "they":

> [4] Replies to comments by D. Domin (1996) on the study by M. B. Nakhleh and J. S. Krajcik in which *they* used concept maps as a method of determining … [*italics added*]

In passage [4], "they", again, seems to avoid the repetition of the names of the authors of the original study, referring the reader back to the AU (author) field to gain the information referent to reply authorship.

In relation to possessive structures (see Figure 5), this investigation yielded the following results: the Saxon genitive was the most frequently used possessive structure accounting for a 2% (f=289) of the total occurrences whereas possessive adjectives accounted for only a 0.7%. The following sample abstract provides examples of the Saxon genitive structures in our corpus:

> [5] Comments on the response of R. P. Honeck and J. G. Temple (see record 1997-04183-002) to the present *authors'* original comment (see record 1997-04183-001) addressing Honeck and *Temple's* (see record 81-39851) article on proverb comprehension and the extended conceptual base and great chain metaphor theories. Issue is taken with *Honeck and Temple's* characterization of the cognitive view and their purported denial of the importance of cultural knowledge in a psychological theory of how proverbs are understood. [*italics added*]

The three Saxon genitive examples in excerpt [5] above seem to be used for a clear expression of author-work relationship, rather than possession; the noun work here might imply cognitive and/or written products or processes.

Possessive Adjective	f	%
my	0	0
your	0	0
his	32	26.229
her	9	7.377
its	9	7.377
our	2	1.639
your	0	0
their	70	57.377

Table 2. PsycLIT RA-C-R abstracts: possessive adjectives.

The frequency distribution of possessive adjectives (see Table 2) showed that the third person plural possessive adjective "their" (f=70; 57.377%) showed the highest frequency, followed by the third person singular possessive adjectives "his" (f=32; 26.229%), "her" (f=9; 7.377%) and "its" (f=9; 7.377%). These results are graphically presented in Figure 5.

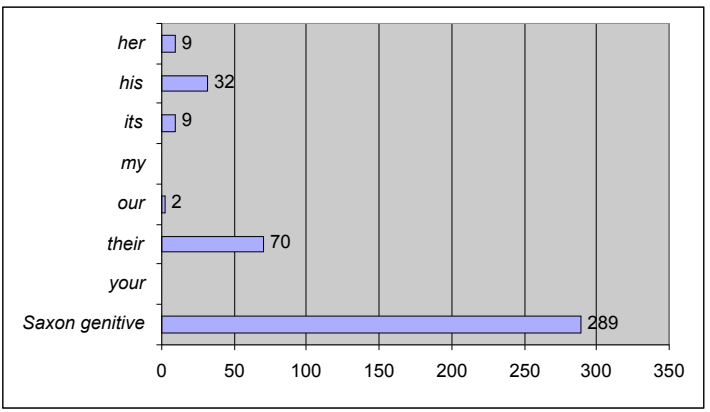

Figure 5. PsycLIT RA-C-R abstracts: possessive adjectives.

The first person plural possessive adjective accounted for only 1.639% of the occurrences. The first and second person singular possessive adjectives showed the lowest scores (0%). These results can also be explained as an avoidance of the "I-author/abstractor" and "you-reader" reference on the part of the writer of the abstract. The first and second persons are avoided by the very nature of scientific text (Weissberg & Buker, 1990).

"Their" seems to have both generic terms and authors' productions as reference terms, as illustrated in examples [6] and [8] below:

> [6] The author notes that not all sexual minority *people* regard these ideas with hostility; and that some feel "liberated" by these ideas, report that the ideas challenge their own sexism and increase *their* sense of solidarity with other kinds of sexual minority people … [*italics added*]

Whereas "his" and "her" have PsycLIT RA-C-R abstracts authors' works as referents together with the common avoidance of sexism in the expression of professions, as illustrated below:

> [7] Replies to a comment by A. Rapaport (1996) on the article by L. Marinoff (see record 81-00423) in response to A. Rapoport's (1964) analysis of the Ace-deuce game. *Marinoff* showed that [...] He re-examines Rapoport's examples to elucidate the flaw in *his* attempt to refute Marinoff's original argument. [*italics added*]

And "its" seems to refer to elements or parts of entities,

> [8] Replies to comment by G. Ryan (16038) on an article by J. M. Wood and L. Wright (see record 83-23242) on children's sexual behaviors and incorporation of base rates in judgment of sexual abuse. The authors refute Ryan's comments by stating that *their article* clearly states **its** purpose, scope, and limitations. [*italics added*]

4.4. Modal verbs

The modal verbs of our corpus (f=92) represented a 0.65% of the total amount of occurrences. As regards to their percent distribution, our data (see Table 3 and Figure 6) showed that "may" (f=23; 25%) was the modal verb most frequently used, closely followed by "can" (f=19; 20.652%).

Modal verb	f	%
can	19	20.652
could	2	2.173
will	10	10.869
would	13	14.130
shall	0	0
should	14	15.217
may	23	25
might	5	5.434
must	6	6.521
ought to	0	0
need	0	0
dare	0	0
used	0	0

Table 3. PsycLIT RA-C-R abstracts: modal verbs.

"May" has been consistently related to hedging (Weissberg & Buker, 1990) and authorial comment (Adams Smith, 1984), whereas "can" has been found to vary across disciplines. "Should" (f=14; 15.217%) and "would" (f=13; 14.130%) which were also quite often used, together with "will" (f=10; 10.869%); "must" (f=6; 6.521%); "might" (f=5; 5.434%); and "could" (f=2; 2.173%) showed the lowest occurrence scores. The modal verbs "shall" –representing this whole group of verbs– and "ought to", and the semimodals "dare", "used" and "need" were absent in our corpus (0%).

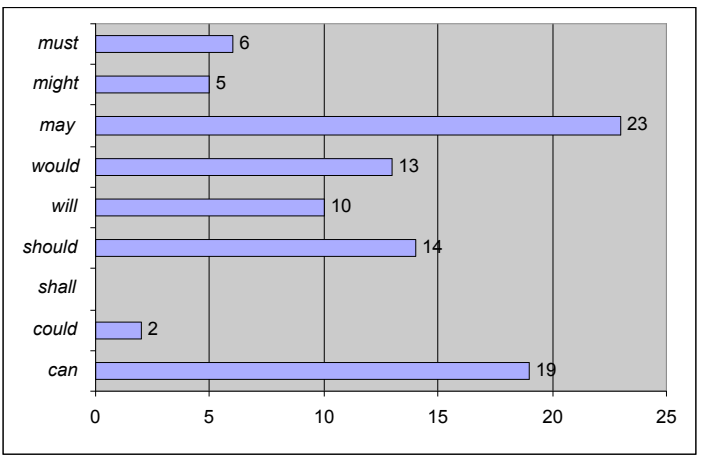

Figure 6. PsycLIT RA-C-R abstracts: modal verbs.

It is frequent to emphasise an attitude of modesty by using modal auxiliaries such as "may", "will", "would", "should" and "could". Although the meaning of these modal auxiliaries differs only slightly from one another, they may be used to highlight a "degree of tentativeness" or an attitude of modesty on behalf of the author, ranging from "SURE" to "TENTATIVE", depending on the modal auxiliary used (Weissberg & Buker, 1990). The presence of epistemic and deontic modality has also been considered an indicator of disciplinary variation in academic English (Piqué-Angordans et al., 2002).

The collocations in Figure 7 illustrate the use of "may", "can", "should", and "would". These contexts suggest that "may" is used in our corpus of PsycLIT RA-C-R abstracts with functions such as the suggestion of further relationships between variables, the indication of presence of sources of

contamination of results, methodological constraints, possible misinterpretation of results, reporting interpretation, possible effects, possibility, function and capabilities, nexi, cause effect relationships, findings and suitability.

should not be assumed to be disease related and	*may*	in fact be drug related
the inclusion of Ss with head injury	*may*	have influenced the results
Continuity between laboratory and field research	*may*	be more apparent than real
The author's article, suggesting than Lewy bodies	*may*	cause a syndrome independent of dement
Conditions under which the use of BPL	*may*	or may not be ethically justified

Figure 7. PsycLIT RA-C-R abstracts: modal verbs (may).

"Can" has also been related to hedging as in the works by Weissberg & Buker (1990). Figure 8 illustrates some of the collocations of can in context.

point that, without MRI coregistration one	*cannot*	conclusively exclude the possibility that be generalized
Unconscious wishes for sexual adventure	*cannot*	differ depending on methodological
It is suggested that stored visual knowledge	*can*	be specific to the form of stimulus
His work on dissociative disorder	*can*	usefully inform those developing the ICD
Propositions about a state of affairs that	*can*	never be perceived directly.

Figure 8. PsycLIT RA-C-R abstracts: modal verbs (can).

In this context, "can" is used for suggesting impossibility of excluding information, evidence or alternative explanations, constraints in the generalisation of results, influencing factors, usefulness of contributions, suggesting relationships, and the impossibility of falsifying hypothesis or observations, together with suggestions for improvement, ways of narrowing the scope of the interpretation, among others.

Instances of "should" and "must" are reproduced in Figure 9. The presence of "should" and "must" suggest the presence not only of epistemic modality, but also of deontic modality in the abstracts of our sample. "Should" seems to be used to indicate and suggest things to do, how things should have been, and to define discussions as part of conditional clauses. "Must" seems to have been used to suggest either appropriate or correct patterns of behaviour and reasoning on the part of the professional psychologist or the researcher. "Will" seems to have been used in the RA-C-R abstracts of our corpus to predict future events in reasoning (concluding), ethics, to indicate future events in the field –repairing lack of acknowledgement, role and importance of things in the future and usefulness of contributions in the field– and to define conditions of research.

Therapists	*should*	employ clinical flexibility
FDG PET in a group of healthy Ss,	*should*	have been cited.
Campbell concerning whether autistic children	*should*	have a trial of naltrextone. Campell did
this can never be a mere intellectual exercise; it	*must*	stem from awareness of lived experience
Results.	will	provide guidance for researchers.

Figure 9. PsycLIT RA-C-R abstracts: modal verbs (should, must, will).

"Would", as shown in Figure 10, might have been used to define hypothetical situations in reasoning procedures, to indicate the topics addressed by the authors of these abstracts, to indicate weaknesses in other authors' productions signalling the effects of different conditions, suggesting alternative ways of acting in the discourse community, among others.

However, if the latter were the case, one	*would*	seek to place the methodological points Maltzman makes…
Zuddas et al concur with Dhossche and Petrides that controlled studies	*would*	be useful to improve therapeutic strategies for young schizophrenic patients…
The authors	*would*	like to underscore the necessity of a gender-inclusive nature for educational goals and pedagogic approaches…
S. Goldstein notes that using Diagnostic and Statistical Manual of Mental Disorders-IV diagnostic criteria	*would*	have been more conclusive if other confounding variables had been controlled…
as it	*would*	have been difficult if not impossible to formulate such an objection at the algorithmic level
He admits that it	*would*	advance the field substantially if controversial articles,… were regularly presented in debate format.

Figure 10. PsycLIT RA-C-R abstracts: modal verbs (would).

4.5. Reported speech structures

Reporting verbs (f=1587) represented an 11.35% of the occurrences of the PsycLIT corpus and a 69% of the verb occurrences. The range of reported speech verbs used more than four times in our sample, together with its frequency (f) distribution, is shown in Table 4. The total number of reporting verbs in the corpus is 1,587. "Reply" accounts for the highest number of occurrences (f=61), which means an 3.968% of the total of reporting verbs; there follow "respond" (f=57), 3.591%; and "argue" (f=37), 2.331%.

verb	f	%	verb	f	%
reply	61	3.968	*compare*	6	0.378
respond	57	3.591	*conclude*	6	0.378
argue	37	2.331	*consider*	6	0.378
suggest	27	1.701	*contend*	6	0.378
comment	22	1.386	*defend*	6	0.378
address	21	1.323	*offer*	6	0.378
discuss	19	1.197	*reduce*	6	0.378
agree	18	1.134	*see*	6	0.378
present	17	1.071	*study*	6	0.378
note	16	1.008	*test*	6	0.378
report	16	1.008	*assess*	5	0.315
show	13	0.819	*attempt*	5	0.315
state	11	0.693	*believe*	5	0.315
assert	10	0.630	*cause*	5	0.315
examine	9	0.567	*challenge*	5	0.315
follow	9	0.567	*demonstrate*	5	0.315
support	9	0.567	*generalise*	5	0.315
propose	8	0.504	*include*	5	0.315
provide	8	0.504	*indicate*	5	0.315
appear	7	0.441	*influence*	5	0.315
clarify	7	0.441	*point out*	5	0.315
represent	7	0.441	*recall*	5	0.315
seem	7	0.441	*relate*	5	0.315
base	6	0.378	*result*	5	0.315

Table 4. PsycLIT RA-C-R abstracts: reporting verbs.

A comparison is called for in this respect with another type of corpus. Hence, a set of 1,132 abstracts drawn from a Medline CD-ROM, amounting to 109,414 words, was also analyzed with the aim of drawing attention to the fact that abstracts from medicine in general, therefore not necessarily in the C-R category as our corpus, would not yield similar results in reference to words we have analysed in our PsycLIT sample, the main concern of our research. See the comparison in Table 5.

	PsycLIT Corpus (13,978 words)		Medline Corpus (109,414 words)	
	f	%	f	%
comment	172	1.231	0	0
reply	90	0.644	4	0.004
respond	57	0.408	4	0.004
study	55	0.393	365	0.332
argue	39	0.279	2	0.002
result	32	0.229	300	0.272
suggest	32	0.229	177	0.160
address	24	0.172	10	0.009
support	13	0.093	51	0.047
propose	10	0.072	29	0.027
provide	9	0.064	65	0.059
appear	7	0.050	44	0.040
clarify	7	0.050	4	0.004

Table 5. PsycLIT vs. Medline abstracts.

From a close look into this table, some data are obviously significant, especially regarding the terms "study", "result" and "suggest", and especially in regard to the first two, which are widely used as nouns rather than verbs. The occurrence of the terms listed is shown graphically in Figure 11. Here these differences clearly stand out, particularly in the almost absolute absence, in the Medline Corpus, of words –verbs and nouns– such as "comment", reply", "respond", and "argue", which appear throughout the PsycLIT Corpus.

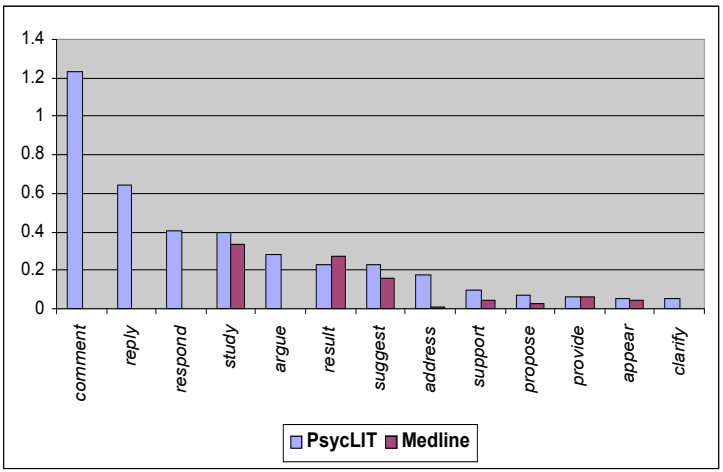

Figure 11. Frequency and percentage use of reporting expressions in PsycLIT and Medline corpora.

Some of the reported verbs listed in Table 4 are shown collocationally in the sample sentences below (Figure 12).

The authors	*assert*	that the psychodynamic explanation.
Although Case	*agrees*	with most of the commentators' points
The author	*addresses*	MCGlashan's theological concerns...
	Replies	to the comment made by S. C Kalischm
	Responds	to the comment made by R. Waltz
P. H. Engebretson and J. Huttenlocher	*commented*	on an article by ...
The author	*defends*	his study by explaining ...

Figure 12. PsycLIT RA-C-R abstracts: reporting verbs in context.

I try to illustrate the occurrence of reporting verbs together with a few sample collocations. Reported speech verbs have also been related to hedging (Varttala, 1999) in specialist texts as an indication of textual precision and interpersonal negative politeness in scholar peer communication.

5. Concluding remarks

An initial sample driven description of PsycLIT RA-C-R abstracts has been attempted in this study, which has served an exploratory purpose, to find out if there exist distinctive features as regards abstract length, number of sentences per abstract, personal pronoun frequency and usage, possessive structures frequency and usage, frequency of word types and reported speech verbs. Results seem to reveal the presence of multiple influences in the abstract features of the sample. Among the typical features of scientific writing studied I have found that impersonality and specificity seem to influence the range of possessive structures and personal pronouns used, together with abstracting journal services features, particular pronoun functions, and sub-genre based features. PsycLIT RA-C-R abstracts also share with other abstracts the presence of hedging, and both deontic and epistemic modality, and a range of variables accounting for their length, sentence length and number of sentences per paragraph. PsycLIT RA-C-R abstracts seem to be distinct in their surface text features as regards to two elements: word types and reported speech structures. Both of them refer to the dialogical content on which the PsycLIT RA-C-R abstracts report.

(Revised paper received November 2007)

References

Adams Smith, D.E. (1984). "Medical discourse: aspects of author's comment". *The ESP Journal* 3: 25-36.

Bazerman, C. (1988). *Shaping Written Knowledge: The Genre and Activity of the Experimental Article in Science.* Madison: University of Wisconsin Press.

Berkenkotter, C. & T.N. Huckin (1995). *Genre Knowledge in Disciplinary Communication: Cognition/Culture/Power*. Hillsdale, NJ: Lawrence Erlbaum.

Bråten, S. (1981). *Time and Dualities in Self-reflective Dialogical Systems. Applied Systems and Cybernetics.* New York: Pergamon Press.

Chubin, D. & E. Hackett (1990). *Peerless Science: Peer Review and U.S. Science Policy.* Albany, NY: SUNY Press.

Ferguson, G. (2001). "If you pop over there: a corpus-based study of conditionals in Medical Discourse". *English for Specific Purposes* 20: 61-82.

Huddleston, R. (1971). *The Sentence in Written English: A Syntactic Study Based on an Analysis of Scientific Texts*. Cambridge: Cambridge University Press.

Hyland, K. (2000). *Disciplinary Discourses.* Harlow: Longman.

Lancaster, F.W. (1991). *Indexing and Abstracting in Theory and Practice.* London: Library Association Publishing Press.

Liebscher, R. & D. Groppe (2003). "Rethinking context availability for concrete and abstract words: a corpus study". In D. Archer, P. Rayson, A. Wilson & T. McEnery (eds.), *Proceedings of the Corpus Linguistics 2003 Conference*, 449-456. Lancaster: Lancaster University.

Love, A. (1993). "Lexico-grammatical features of geology textbooks: process and product revisited". *English for Specific Purposes* 12: 197-218.

MacDonald, S. P. (1990). "The literary argument and its discursive conventions". In W. Nash (ed.), 31-62.

Merton, R.K. & S.H. Zuckerman (1973). "Institutionalized pat-

terns of evaluation in science". In R.K. Merton (ed.), *The Sociology of Science: Theoretical and Empirical Investigations*, 460-496. Chicago: University of Chicago Press.

Motta-Roth, D. (1998). "Discourse analysis and academic book reviews: A study of text and disciplinary cultures". In I. Fortanet, S. Posteguillo, J.C. Palmer & J.F. Coll (eds.), *Genre Studies in English for Academic Purposes*, 29-58. Castelló: Universitat Jaume I.

Myers, G. (1990). *Writing Biology: Texts in the Social Construction of Scientific Knowledge*. Madison: University of Wisconsin Press.

Nash, W. (ed.) (1990). *The Writing Scholar. Studies in Academic Discourse*. Newbury Park: Sage Publications.

Piqué-Angordans, J. & M.J. Coperías Aguilar (1999). "Verb tense in essay writing: frequency of use in health sciences". In S. Barrueco, E. Hernández & L. Sierra (eds.), *Lenguas para Fines Específicos-VI. Investigación y enseñanza*, 245-251. Alcalá de Henares: Servicio de Publicaciones de la Universidad de Alcalá.

Piqué-Angordans, J. & J.V. Andreu-Besó (2000). "A textual perspective of scientific articles: Pattems and moves". In M. Ruane & D.P. Ó Baoill (eds.), *Integrating Theory and Practice in LSP and LAP*, 57-70. Dublin: University College Dublin & IRAAL.

Piqué-Angordans, J. & S. Posteguillo (2006). "Peer positive and negative assessment in Medical English written genres". In M. Gotti & F. Salager-Meyer (eds.), *Advances in Medical Discourse Analysis: Oral and Written Contexts*, 386-406. Bern: Peter Lang.

Piqué-Angordans, J., S. Posteguillo & J.V. Andreu-Besó (2002). "Epistemic and deontic modality: a linguistic indicator of disciplinary variation in academic English". *LSP & Professional Communication* 2: 49-65.

Scott, M. (1996). *WordSmith Tools*. Oxford: Oxford University Press.

Sionis, C. (1995). "Communication strategies in the writing of scientific research articles by non-native users of English". *English for Specific Purposes* 14: 99-113.

Swales, J. M. (1990). *Genre Analysis: English in Academic and Research Settings.* Cambridge: Cambridge University Press.

Swales, J. M. (2004). "Then and now: A reconsideration of the first corpus of scientific English". *Ibérica* 8: 5-21.

Swales, J. M. (2006). "Corpus linguistics in English for Academic Purposes". In E. Arnó Macià, A. Soler Cervera & C. Rueda Ramos (eds.), *Information Technology in Languages for Specific Purposes. Issues and Prospects*, 19-33. New York: Springer.

Upton, T.A. & U. Connor (2001). "Using computerized corpus analyses to investigate textlinguistic discourse moves of a genre". *English for Specific Purposes* 20: 313-329.

Van Peer, W. (1990) "Writing as an institutional practice". In W. Nash (ed.), 192-204.

Varttala, T. (1999). "Remarks on the communicative functions of hedging in popular scientific and specialist research articles in medicine". *English for Specific Purposes* 18: 177-200.

Weissberg, R. & S. Buker (1990). *Writing Up Research. Experimental Research Report Writing for Students of English*. Englewood Cliffs, NJ: Prentice Hall Regents.

Ingrid García Østbye, while working as an English instructor at the Escuela Oficial de Idiomas in Quart de Poblet (Valencia), is researching towards her doctoral dissertation in the area of discourse analysis. Her main interests are teaching, translation studies and discourse analysis, specifically in the field of psychology and health sciences.

NOTES

[1] PsycLIT, currently called PsycInfo, is one of the scientific databases in psychology; many universities are connected to PsycLIT databases on CD-ROM and online. Its journal coverage, spanning 1887 to present, includes international material selected from more than 1,300 periodicals written in over 25 languages.

[2] The search was carried out at one of the search points (an Apple MacIntosh) of the library in the Facultat de Psicologia at the Universitat de València.

[3] In this sense Swales (2006) refers to the need for small corpora.

[4] Although only 22 of these 172 times are considered as reporting verbs (see Table 1), for comparison purposes, "comment/s" has also been included as a noun.

Appendix: A PsycLIT RA-C-R abstract.

PsycLIT 1996-3/98
Record 1

ABSTRACT: In this brief comment on J. M. Wood, M. T. Nezworski, and W. J. Stejskal's (1997) response to his article (Meyer, 1997a), the author documents how J. M. Wood et al. continue to make allegations based on a limited subset of the available literature. He also points out specifically how their criticisms regarding kappa, test-retest reliability, true score theory, score aggregation, and his meta-analysis are incorrect. He concludes that these new errors provide additional reasons to be cautious about the conclusions proffered in their other articles on the Rorschach. ((c) 1998 APA/PsycINFO, all rights reserved)(journal abstract)
KEY PHRASE: meta analysis of interrater reliability and evaluation of reliability problems reported by J. M. Wood et al in Rorschach Comprehensive System, commentary reply
MAJOR DESCRIPTORS: *Interrater-Reliability; *Meta-Analysis; *Rorschach-Test; *Test-Interpretation; *Test-Reliability
MINOR DESCRIPTORS: Professional-Criticism-Reply