

*The Good Me or the Bad Me?  
Identity and Evaluation in Research Article Abstracts*

Numerosi linguisti, tra cui Bhatia 1993 e 1997, hanno ampiamente dimostrato che il genere dell'abstract, inteso come testo accademico, non risulta essere più oggettivo di altri generi testuali ma semplicemente è più efficace a nascondere la voce individuale dell'autore.

L'obiettivo di questo articolo è quello di analizzare la modalità con cui gli autori valutano e presentano i loro risultati ai loro colleghi e pari con il fine ultimo, da un lato, di difendere la loro identità e dall'altro, di costituire una comunità discorsiva. Dal momento che la ricerca si focalizza più ampiamente sul discorso valutativo (Aijmer 2005, Mauranen 2004, Stubbs 2001, e Swales 2004) ed identitario, verranno presi in considerazione elementi lessicali definiti nello studio come *research process words (RPWs)*. Attraverso un'attenta analisi delle *collocations* delle *RPWs*, intendiamo offrire un mezzo di analisi del linguaggio valutativo laddove esso si realizza in presenza di parole specifiche. L'analisi di specifiche costruzioni ed elementi lessicali si avvale anche delle frequenze di distribuzione comparativa. I risultati del presente studio supportano l'ipotesi che il linguaggio valutativo incorpora un'interazione tra autore e lettore indipendentemente dalla disciplina in oggetto. Precise scelte lessicali e parole specifiche appaiono essere più frequenti di altre. Tuttavia resta inteso che l'analisi testuale e l'identificazione del linguaggio valutativo non sono processi immediati anche in studi accademici che si avvalgono di tecniche analitiche ed automatiche.

*1. Introduction*

Within their discourse community, academics use language to acknowledge, construct and negotiate social relations through various institutionally recognised genres, such as lectures, papers presented at conferences, research articles and monographs. In this paper, we discuss the way academic identity is realised through interactions between writers and readers in one specific genre, namely the research article abstract (RAA). In RAAs authors seek to offer a credible representation

of themselves and their work by claiming solidarity with their discourse community and readers, evaluating material and acknowledging alternative views. In this perspective, linguistic choices are central to the building of a convincing argumentation. It is therefore our aim to highlight some of the linguistic choices that authors adopt in two scholarly journals (*The International Journal of Primatology* and *Mathematics and Computers in Simulation*) in order to position themselves within their own discourse community while positively evaluating their research at the same time.

### *1.1. Research article abstracts and evaluation*

The RAA is an evaluative genre by definition (Bhatia 1993; Swales 1990). As Mauranen (2004: 207) suggests, “evaluation is an interesting phenomenon, being a central aspect of what academics do. We do not get published if we only present results, we also have to evaluate”. Hunston (1993, 1994) has described how experimental research papers are largely evaluative. She claims that academic articles belong to the value system of ‘good research’, which means that even if the markers of attitudinal lexis are missing, the writers’ attitude to the value of their research is clear: the aim is to present something positive and appealing to the target discourse community. RAAs have aroused great interest due to the important role they fulfil in academia and have been investigated in various ways by different authors. Bhatia (1993: 78) defines the RAA as “a description or factual summary of the much longer report, [...] meant to give the reader an exact and concise knowledge of the full article”. Among others, Salager-Meyer (1990), Gibson (1993), and Dos Santos (1996) regard RAAs as an independent genre whose explicit function is to provide information about the content of a research paper, indicating to readers that the full text merits further attention. Such studies place special emphasis on the evaluative aspects of the genre. In this perspective, Martín-Martín (2005: 5) suggests that:

In the process of publishing the results of research, abstracts constitute, after the paper’s title, the readers’ first encounter with the text, and it is here that writers have to show they have mastered the conventions (the

textual organization and other rhetorical practices) that are favoured by the members of a specific disciplinary group.

Abstracts therefore play various roles: first they help the reader to ascertain the paper's purpose, then they provide the reader with a preliminary overview of the research and, in some cases, may help to recall its basic content. Publishing, as observed by Swales (1990), is a way to join a discourse community, so that authors need to persuade their audience of what they say, to press their point of view by means of an articulate, competent use of language. The standard style of academic writing suggests that it should be objective not attitudinal, and that the text should be entirely free of personal judgements. This, however, does not happen very often because one of the chief functions of scientific research articles is to persuade the reader of the validity of the writer's claims, and in order to accomplish this purpose, the work of the writers and of other researchers is constantly evaluated along the text distribution (Hunston 1993, 1994). Academic writing is as rhetorical as any other type of discourse, no matter how technical and apparently detached it might appear, as its discourse is always designed to convince readers of the reliability of its claims. Hunston therefore rejects the widespread idea that "evaluation is personal and scientific writing impersonal" (1983: 58).

Evaluation is not the only key feature of RAAs, however; a crucial role may also be played by identity. Hyland (2000: 63) states that RAAs are "a rich source of interactional features that allow us to see how individuals work to position themselves within their communities". Other studies (e.g. Myers 1990) have shown that before a research paper is published, a great deal of negotiation on the final version to be published goes on between authors, editors and referees. Researchers have to argue their case in front of the bar of the scientific community before their work can be taken up and accepted. Myers (1985) argues that the tension inherent in the publication of any research article makes negotiation between the writer and the potential audience essential: the researcher tries to show that s/he deserves credit for something new, while on the other hand the editors try to relate the article's claims to a body of existing knowledge. Thus the focus shifts from the individual researcher to the entire research community, from authorial identity to

collective identity. In this perspective, there is a direct equation between what is said and how the author's identity is represented, as "publication can be seen as documentary evidence that the writer qualifies for membership in the target discourse community" (Swales 1990: 7).

## 2. *Texts and methodology*

The present study is based on an Abstract Corpus (ABS) recently assembled by the author (cf. Cava 2008) which comprises 1,035 RAAs from two international scientific journals: *The International Journal of Primatology* (360 texts) and *Mathematics and Computers in Simulation* (675 texts), for a total of around 200,000 words. The time span taken into consideration is 2000-2005. *The International Journal of Primatology* publishes laboratory and field studies related to anatomy, ethology, cognition, ecology, conservation, genetics, evolution and physiology in primates. *Mathematics and Computers in Simulation* publishes articles on specific applications of modelling and simulation in science and engineering, relevant to applied mathematics, the general philosophy of systems simulation and their impact on disciplinary and interdisciplinary research.

According to the American National Standard for Writing Abstracts (ANSI 1979: 1), "an abstract is an abbreviated, accurate representation of the contents of a document, preferably prepared by its author(s) for publication with it". In the present investigation, the RAAs consist of a title, the body of the text, and keywords. Information about authors and their institutions were deleted because not relevant to the purpose of the present study. In *The International Journal of Primatology*, guidelines are provided as to the length of abstracts (up to 250 words) and the use of four or five keywords. On the other hand, no specific details are provided in *Mathematics and Computers in Simulation*. In neither case do editorial guidelines provide clear indications for writing and structuring RAAs.

In *The International Journal of Primatology*, the average length of RAAs is between 200 and 250 words, but one abstract has 377 words and another that only 88. These two borderline cases show that text length requirements are not always observed literally. In *Mathematics*

*and Computers in Simulation* the average length is shorter (about 150 words), with a minimum of 49 and a maximum of 266 words. The RAAs in both journals can be defined as informative, because they provide information about the contents of the article, its key facts and conclusions.

Following the principles of corpus linguistics, our analysis compared the linguistic features of the texts under investigation against a reference corpus. We opted for the Corpus of Contemporary American English (Davies 2008), which contains a large section of academic writing and offers a balanced sample of contemporary American English. It contains more than 385 million words of text, 20 million of which from 1990-2008, and is equally divided between speech, fiction, popular magazines, newspapers and academic texts. The corpus also allows easy searches and comparisons of the frequency of words, phrases and grammatical constructions by genre and over time. We therefore limited our investigation to a single section of COCA, the Academic Journals subcorpus, which consists of more than 73 million words, representing nearly 100 different peer-reviewed journals covering the entire range of the Library of Congress classification system.

### *2.1. Preliminary findings*

To conduct a corpus-based analysis of discursual features such as identity value and evaluation, words cannot be analysed in isolation but long stretches of text need to be taken into account. Evaluation is not only a lexical phenomenon but operates across the text as a means of cohesion, being a multifunctional phenomenon deeply connected to the author's identity: it shows a "context-dependent polysemous functionality" (White 2001: 18). At the same time evaluation can be used to express the writer's opinion, or to construct relations between the writer and the reader (Hunston / Thompson 2000; Thompson / Ye 1991).

The point of entry in our analysis focuses on the 'lexical items' (Sinclair 1996) that operate as *research process words* (henceforth RPWs) in RAAs identified in an earlier study (Cava 2008). These are related to the research process, that is they belong in Halliday's (1994) terms to the material process of doing, although the logical elements of actor, process and goal are in a way present only in the words

themselves. In Halliday's example *the lion caught the tourist*, the *lion* is the actor, *caught* is the process and the *tourist* is the goal. The words identified for investigation are *analysis/es*, *data*, *evidence/s*, *findings*, *investigation/s*, *methods*, *methodology/ies*, *paper/s*, *procedure/s*, *research/s*, *result/s*, *study/ies*, and *theory/ies*. These all have the same actor (the researcher) and encapsulate the action. The research process implies that the researcher analyses and investigates data (that were previously collected) by a viable method for a specific purpose; then, s/he will obtain evidence and findings that make up a theory which, eventually, will allow the researcher to describe the results in a paper. Accordingly, these can be defined as *research process words* because they emphasise the process that underlies RAAs. The following table lists in alphabetical order the frequency of RPWs in the ABS corpus.

<i>RPW</i>	<i>Frequency</i>
Analysis	275
Analyses	52
Data	375
Evidence	82
Evidences	2
Finding	32
Findings	53
Investigation	34
Investigations	10
Method	537
Methods	248
Methodology	19
Methodologies	1
Paper	346
Papers	6
Procedure	33
Procedures	17
Research	69
Researches	0

Result	62
Results	386
Study	240
Studies	111
Theory	110
Theories	9

Table 1. Frequency of research process words in ABS.

Some of these words appear among the top 100 content words in the ABS wordlist. This result confirms the assumption that a specific genre, in this case the RAA, is characterised by specific lexical choices. However, the raw frequencies of RPWs in ABS do not provide us with enough information: a more complete picture should emerge by comparing the most frequent RPWs in ABS and in the reference corpus (COCA). In order to carry out this comparison, the frequencies in both corpora were normalised per 1,000 words.

<i>RPW</i>	<i>ABS</i>	<i>COCA</i>
method	3.00	0.17
results	2.16	0.42
data	2.10	0.62
paper	1.94	0.16
study	1.34	0.83

Table 2. Most frequent research process words in ABS and COCA (normalised per 1,000 words).

The normalised figures in Table 2 clearly show that the relative frequency of RPWs is far higher in RAAs, compared to research articles as a whole (COCA). This confirms our initial hypothesis that the focus on the research process is one of the key features of RAAs.

### 3. Analysis and results

The preliminary findings yielded by the comparison of normalised frequencies in ABS and COCA required further investigation in order to account for the role of RPWs. Running concordances of the RPWs can tell us something about the ABS corpus and its peculiarity compared to COCA. Due to the difference in size between the two corpora, collocates were compared based on their statistically significant Mutual Information (MI) score rather than mere frequency. As suggested by Scott (2004), the MI score relates one word to another, taking into account not just the most frequent words found close to the word in question, but also how often each word occurs elsewhere in the text. Therefore the MI score identifies those collocates that have a strong co-occurrence with the node word.

Collocates identified with this procedure were grouped into 3 main categories: those related to a procedural aspect; those related to evaluation; and those related to identity. Procedural words include the verbs *solve* and all its lemmatised forms (e.g. *solution*, *solving*, *solves*) and *apply* with all RPWs, since the procedural aspect is related to function and operation performed by the researcher as described in the article. On the other hand, words that can be considered evaluative *per se* are adjectives such as *effective*, *new*, *accurate* and *relevant* or verbs like *provide* and *corroborate*; evaluative adjectives are generally polar (e.g. *good* or *bad*), so that identifying evaluation is relatively straightforward. Finally, words grouped in the category of identity include personal pronouns *I* and *we*, the possessive adjectives *my* and *our* and the deictic *this*, which encapsulate the presence of the author and his/her own discourse community.

Relevant collocates of the *method*, which is the most frequent RPW in ABS, are listed in the table below:

Category	Collocates in ABS	Collocates in COCA
Identity	this (5.92)	this (0.98)
	we (3.65)	our (0.01)
Evaluation	new (3.86)	effective (1.93)
		new (0.08)

Procedure	solving (8.97)	procedure (2.21)
	solution (8.02)	used (1.92)
	solve (6.53)	developed (1.67)
	developed (8.38)	analysis (1.45)
	applied (8.11)	data (1.21)
	proposed (4.66)	study (1.18)
	paper (6.55)	research (0.78)
		results (0.67)

Table 3. Collocates of *method* in ABS and COCA (MI score in brackets).

Due to space constraints, we concentrate on identity collocates, which are more explicit in signalling authorial presence. MI scores are higher in the ABS corpus, where language is used for a more focused communicative purpose. The most recurrent collocates are related to procedure, especially the lemmatised forms of *solve* (*solve*, *solving*, *solution*). The strong presence of these collocates in ABS and their absence in COCA highlights the researchers' role and their effort to operate within the problem-solution pattern.

Concgrams were then generated to investigate lexical patterns in the corpus. As defined by Cheng *et al.* (2006: 414), a concgram "is all of the permutations of constituency variation and positional variation generated by the association of two or more words". This means that the associated words comprising a particular concgram may be the source of a number of 'collocational patterns'. The following concgram of *we* + *method* shows all co-occurrences of the two terms, arranged in such a way as to highlight their usage in the ABS corpus:

algorithm.</p> <p>An important feature of the method **we** present lies in its valid dissipator</keywords> <p>By using the Fourier method **we** study the stability of a  
 Applying the Von-Neumann stability analysis method **we** show that the proposed me  
 by Kusuoka to finance problems. By using *this* method, **we** achieve 6500 times faste  
 of 297 agonistic con- flicts with the PC-MC method: **we** observed focal individua  
 Phys. 174 (2001) 946]. As verification of the method, **we** tabulate the h-condition  
*this* paper a one parameter predictor-corrector method, which **we** call it A-EBDF, is  
 proximates the Signorini condition. **We** use the method of lines to obtain numerical  
 these overheads are prohibitive. **We** present a method for solving the mesh partiti  
 educed phase spaces</keywords> <p>**We** present a method to calculate formal symmetri  
 Using the local Lagrangian form **we** extend the method of Marsden, Patrick and Schk

e infinite-dimensional case, we use Galerkin's method to reduce the space. Numeric stochastic approach. First, we use Monte Carlo method to sample and to build much

<p>In *this* paper, we use an algebraic method to compute the discrepancy</keywords> <p>We present a particle method for solving initial-value pr dynamics</keywords> <p>We present a numerical method that allows a formation of c A-BDF and EBDF methods we propose a multistep method whose region of absolute sta

<p>In *this* paper, we present a numerical method for the computation of surfa analysis method we show that the proposed method is unconditionally stable. B secondary forest. We used a marked-nest census method to examine seasonal changes discrete method. We show that accuracy of the method for quadratic functions impr cost function, we study the stability of *this* method and present some numericals

In Section 1, we give a brief sketch of *this* method. In Section 2, we will expla method. We establish convergence of the method and prove error estimates in problems. We discuss two versions of *this* method: (a) Finite volume discontin computation, we present a general and unified method for investigating the genera

In *this* paper we introduce a finite difference method for a numerical simulation o and thus, we call it the 'moving index' (MI) method. We use the so-called linear primates, we developed a 4-step noncorrection- method -type finger maze (4FM) based cond part we will outline a recently developed method that is based on conformal m

we will test the accuracy of the proposed method.</p> </id> <id="69"> <head>

We employ a version of the finite element method to discretize the space of s

We find the set of conditions for which each method is more advantageous than th

We use the fuzzy pattern matching (FPM) as a method of classification and the tr

we will illustrate the feasibility of *this* new method.</p> </id> <id="98"> <head>

We prove the convergence of the secant modules method to the exact solution. The p

we can include the incremental learning in *this* method, and we compare the obtained

we propound the use of a Monte Carlo simulation method based on the Kolmogorov and

### Concgram 1. Occurrences of *we + method* in ABS.

Recurrent patterns are *we use the method* or *we present the method*. Authorial identity is relevant as long as it is useful to the purpose of the research paper itself. Acknowledging the use of a specific algorithm or DNA sequence technique allows authors to place themselves in an authoritative position, establishing their own territory and own credentials. The text follows a sequence of deductive reasoning whose conclusions are derived from two premises – as in a syllogism. In presenting his/her research, the author also argues that it is worth publishing, since academic discourse “is *designed* to persuade readers of the objectivity of its methods and the correctness of its findings” (Drew 2004: 217).

A strong presence of the identity markers in the ABS corpus as compared to COCA is noticeable among the collocates co-occurring with the RPW *results*:

Category	Collocates in ABS	Collocates in COCA
Identity	these (7.12)	these (1.47)
	we (5.02)	
	our (4.95)	
	this (1.49)	
Evaluation	experimental (5.54)	preliminary (3.06)
	well (4.88)	experimental (1.96)
	positive (1.60)	
Procedure	show (9.21)	indicate (3.29)
	indicate (8.04)	yielded (3.24)
	present (7.95)	interpreting (3.07)
	presented (7.25)	indicated (3.05)
	simulation (7.16)	suggest (3.02)
	paper (5.66)	
	suggest (5.42)	

Table 4. Collocates of *results* in ABS and COCA (MI score in brackets).

In the Procedure group, the main difference between corpora lies in the presence of two procedural collocates (*present* and *presented*) that signal more explicitly the author's presence and academic identity as part of the research process. Once again there is a stronger presence of identity elements in ABS than in COCA.

The next concgram shows the not easily predictable co-occurrence of *results* with the demonstrative determiner *this*:

based on craniodental variation. Results of **this** investigation suggest patterns processing raw data, plugging the results of **this** processing into theoretical model to infants than mothers did. The results of **this** study emphasize the existence of a capture point coordinates. The results of **this** task are used in the tuning of not otherwise essential habitat. Results of **this** study have implications for important experimental observations. The results reported in **this** paper are a powerful step. Simulations confirm **this** and results are compared with well-known numerical phism are indicated by **this** test. Results of the discriminant function analyses (onent mixture theory. **This** theory results in a set of coupled non-linear partial /keywords> <p>**This** paper outlines results of a sensitivity analysis on a model de /keywords> <p>**This** paper analyses results from an investigation into the determin ywords> <p>**This** paper reports the results of experimental and model studies of th

rds> <p>In **this** paper, we present results concerning the far field pattern genera  
 plasma. In **this** paper, we present results of a numerical simulation of semiconduc  
 n models. **This** paper presents the results of experiments designed to track humans  
 ords> <p>**This** paper presents some results from a numerical model of the wind and  
 parties. **This** finding mirrors the results concerning reconciliation in spectacl  
 ed sharing **this** property. Numerical results illustrate the usefulness of these new  
 roblem. **This** paper gives explicit results that simplify the implementation of the  
 <p>In **this** paper, computational results for the finite time case are presented.  
 im of **this** report is to give some results of numerical experiments and discuss it  
 ed on **this**, along with well known results on local existence and uniqueness of so  
 p>In **this** paper, we review recent results concerning stochastic models for coagul  
 hes. **This** paper reviews stability results of several velocity-pressure pairs with  
 p>**This** is a short survey of known results about elimination of quantifiers over n  
 ect was independent from age. The results indicate that the possibility of techni

Concgram 2. Occurrences of *this* + results in ABS.

These co-occurrences reflect the frequent reference to the research  
 process (*this investigation; this study*) and more specifically to the  
*paper* itself. The clause *In this paper we present results* signals authorial  
 identity through the personal pronoun *we*, but at the same time the  
 authors' identity is reinforced by the cluster *this paper*, which points  
 directly to the author. What the author writes is his/her product in a way  
 that establishes a position within the community: as Swales (1990:7)  
 remarks, "publication can be seen as documentary evidence that the  
 writer qualifies for membership in the target discourse community".

The analysis of the collocates of the RPW *data* in the two corpora  
 provides rather unexpected results:

Category	Collocates in ABS	Collocates in COCA
Identity	our (5.97)	
	this (5.83)	
	these (4.54)	
	I (3.81)	
	the (3.13)	
	we (1.34)	
Evaluation	provide (7.80)	accurate (1.58)
	experimental (7.75)	relevant (1.53)
	new (6.42)	reveal (1.53)

Procedure	collected (6.38)	collected (4.06)
	based (3.39)	collect (3.78)
		collection (3.62)
		collecting (3.58)
		gathered (3.42)
		analyzed (3.25)
		gather (3.14)

Table 5. Collocates of *data* in ABS and COCA (MI score in brackets).

Procedural items appear to be less prominent among the collocates of *data*, while the identity lexis is more evident. No collocates related to identity occurred in the reference corpus, as identified by MI score. The explanation may be that *data* are usually chosen and collected by the researcher himself; so that the analysis of these ad hoc data, implies specific choice made a priori. In a similar way when a researcher intends to test an hypothesis data are collected keeping in mind this premise. The relatively high number of collocates such as personal pronouns (*I* and *we*), possessive adjectives (*our*) and the demonstrative determiners (*this* and *these*) reinforces our hypothesis that building identity value is a relevant issue that emerges from the comparison of RAAs with scientific writing in general.

In the next congram of the pattern *we collected data*, the most common cluster or 3-gram among collocates of *data* clearly shows how authorial identity is explicitly referred to when *data* are present and discussed in RAAs:

group. **We collected** behavioral and hormonal data during 7 mo from the reproduct  
 Eritrea, **we collected** detailed demographic data on six bands at four sites in  
 March 2000. **We collected** 92 h of behavioral data in 76 sessions. The infants we  
 Forest</head> <p>**We** systematically **collected** data on feeding behavior for one gr  
 emur fulvus rufus.**We collected** scent-marking data on adult male Lemur catta at B  
 northeastern Brazil. **We collected** behavioral data as all day follows, once a wee  
 ypoxanthus</head> <p>**We collected** systematic data on the home range and day rang  
 search Center</head> <p>**We collected** nesting data from 512 fresh nest sites, inc  
 nd other visible injuries. **We also collected** data on the demography, biogeograph  
 of them in 1996-1997 and 2003. **We collected** data on vegetation fragments and thi  
 ui) on Yakushima Island, Japan. **We collected** data on their activity budgets, qua  
 e Chaco. Between 1997 and 2000, **we collected** data to evaluate the relationship b

onobos (*Pan paniscus*) We collected data on parasitic prevalence and leaternity to 13 candidate males. We collected data for 19 females that had given relation to forest seasonality. We collected data over 12 mo in lowland dipterocidly behavior toward mothers. We collected data by observing all individuals w with a very long PC duration. We collected data on 2 captive groups of ring-ta to social or breeding status. We collected data during a 12-mo study on 2 grou f postcopulatory female choice. We collected data on female sexual swellings, se nd Daytime Activities We collected data on diet and activity budget in nongjia Nature Reserve, China. We collected data during 2 winters from 1998 to grouping patterns than elsewhere. We used data collected on food supply and p

### Concgram 3. Occurrences of *we + collected + data* in ABS.

Of the five most frequent RPWs in the ABS corpus, we have analysed the first three (*method*, *results* and *data*). The other two (*paper* and *study*) are likely to follow similar patterns to those presented so far, regardless of differences in the way they are used in abstracts and in academic writing in general. In either case, the presence of the author is more explicitly visible. Interestingly, *paper* is often used in the ABS corpus to present *new* findings, as the concgram below clearly illustrates:

ors to solve this problem. In this paper, a **new** fast implementation in the freque ing results. In this context, this paper presents a **new** method based on the resca gation process. We present in this paper a **new** approach that uses visual informat ow principle In this paper, a **new** approach for robust fault detecti r prediction In this paper, a **new** method is presented that offers e r control purposes. In this paper, a **new** model structure for the simulatio erical stability problems. In this paper, a **new** type of algorithm to solve the di hich are discrete in time. In this paper, several **new** theorems on the stability o tteny problem is proposed in this paper. In this **new** method, the original data i global sensitivity analysis. This paper presents a **new** version, incorporating sh on the robustness criterion, this paper provides a **new** way to deal with the stab amily of scrambled sequences. This paper presents a **new** algorithm for finding an the choice of generator set. This paper describes a **new** method for the construct dge; Simulation This paper describes a **new** specific blackboard expe es and up to fourth order. In this paper, we present **new** optimal fourth-order SSP chastic process This paper considers a **new** class of time series mod utions which are described in this paper. Firstly, a **new** graphical user interface el computing In this paper, we present a **new** approach for the paral The purpose of this paper is to propose a **new** method for blind equ Renormalization This paper is devoted to the presentation of **new** me 0, the **new** method provided in this paper is proved effective and practical in the

### Concgram 4. Occurrences of *this + paper + new* in ABS.

This investigation of RAAs suggests that the lexis of identity and evaluation is more prominent in our corpus rather than in reference corpus. As argued by Hyland, the use of personal subjects “in academic writing reveals overt acceptance of personal responsibility” for the claims being made (1998: 181). To test our findings we therefore decided to compare the normalised frequency of the personal pronoun *we* in the two corpora: *we* occurs 7.2 times per 1,000 words in ABS and only 2.19 times in the reference corpus. In ABS not only does *we* occur three times more often but also the verbs collocating with it show greater commitment on the part of authors. Table 6 lists the most relevant of these verbs in the two corpora: in ABS the high score of *prove* and *demonstrate* – which can be defined as mental verbs with a controlling function and “an important device to express stance” (Biber *et al.* 1999: 666) – reflect a strong commitment of authors to the importance of the research reported in the abstract.

<i>ABS</i>	<i>MI score</i>	<i>COCA</i>	<i>MI score</i>
prove	12.38	conclude	3.02
conclude	9.70	infer	3.00
demonstrate	7.79	speculate	2.99
show	7.34	observe	2.49
suggest	6.04	discover	2.42

Table 6. Collocates of *we* in ABS and COCA.

Collocates occurring only in the reference corpus are *infer*, *speculate*, *observe* and *discover*. These can be grouped as cognition and perception verbs because they deal with the realm of ‘discovery and description’. These verbs are very frequent in academic prose because as verbs they do imply a process but moreover it is strongly supported by reason and by deduction. The different collocational patterns are further evidence of the importance attached to RAAs in the identity building process carried out in academic discourse.

#### 4. Conclusions

In this paper, we have analysed the way authors offer a positive representation of themselves and their work by means of a collocational analysis of specific lexical items, i.e. research process words, in a corpus of research article abstracts published by two international journals, *The International Journal of Primatology* and *Mathematics and Computers in Simulation*. RAAs were chosen both because they are considered a highly 'evaluative' genre (Bhatia 1993 and Swales 1990) and because they are used by authors to position themselves within their discourse community (Hyland 2000). The comparison of the most frequent RPWs in the ABS corpus and in the Academic Journals sub-corpus of the *Corpus of Contemporary American English* has shown how their relative frequency is extremely higher in the RAAs, confirming our hypothesis that the focus on the research process is one of the key features of abstracts.

The analysis of collocates and congrams of the most frequent RPWs has revealed a significant co-occurrence with terms specifically related to the presence of the researcher, when the research process is discussed. The strong presence of these collocates in the ABS corpus elicits the researchers' role, generally portrayed in a positive way. Identity value and evaluative language appear to be more evident in the ABS corpus rather than in the reference corpus confirming our initial hypothesis that abstracts as a genre are the site adopted by authors to place themselves and also to place the readers within their own discourse community in a position from which to strive for positive evaluation.

#### References

- ANSI (American National Standards Institute), 1979, *American National Standard for Writing Abstracts* (Tech. Rep. No. Z39.14-1997), New York, ANSI.
- Bhatia, Vijay K., 1993, *Analysing Genre: Language Use in Professional Settings*, London, Longman.
- Biber, Douglas / Johansson, Stig / Leech, Geoffrey / Conrad, Susan / Finegan, Edward, 1999, *Longman Grammar of Spoken and Written English*, Harlow, Longman.

- Cava, Amelia Maria, 2008, *Good vs Bad in Research Article Abstracts a corpus-based analysis*, Unpublished PhD dissertation, Naples, University of Naples - Federico II.
- Cheng, Winnie / Greaves, Chris / Warren, Martin, 2006, "From n-Gram to Skipgram to Congram". *International Journal of Corpus Linguistics* 11/4: 411-33.
- Cleveland, Donald B., 1983, *Introduction to Indexing and Abstracting*, Englewood, Libraries Unlimited.
- Davies, Mark, 2008, *Corpus of Contemporary American English (COCA)*, Birgham Young University. Available at <http://www.americancorpus.org> [October 15, 2008].
- Dos Santos, Mauro B., 1996, "The Textual Organization of Research Paper Abstracts in Applied Linguistics". *Text* 16/4: 481-499.
- Drew, Paul, 2004, "Integrating qualitative Analysis of Evaluative Discourse with the Quantitative Approach of Corpus Linguistics". In: Tognini-Bonelli, Elena / Del Lungo Camiciotti, Gabriella (eds.), *Strategies in Academic Discourse*. Amsterdam, John Benjamins Publishing: 217-229.
- Gibson, Thomas R., 1993, *Towards a Discourse Theory of Abstracts and Abstracting*, Vol. 1, Nottingham, University of Nottingham.
- Halliday, Michael A.K., 1994, *An Introduction to Functional Grammar*, London, Edward Arnold.
- Hoey, Michael, 1983, *On the Surface of Discourse*, London, Allen & Unwin.
- Hoey, Michael, 2001, *Textual Interaction: An Introduction to Written Discourse Analysis*, London, Routledge.
- Hunston, Susan, 1983, "Text in World and World in Text: Goals and Models of Scientific Writing". *Nottingham Linguistic Circular* 14: 25-40.
- Hunston, Susan, 1993, "Evaluation and Ideology in Scientific Writing". In: Ghadessy, Mohsen (ed.), *Register Analysis: Theory and Practice*, London, Pinter: 57-73.
- Hunston, Susan, 1994, "Evaluation and Organization in a Sample of Written Academic Discourse". In: Coulthard, Michael (ed.), *Advances in Written Text Analysis*, London, Routledge: 191-218.
- Hunston, Susan / Thompson, Geoff (eds.), 2000, *Evaluation in Text: Authorial Stance and the Construction of Discourse*, Oxford, Oxford University Press.
- Hyland, Ken, 1998, *Hedging in Scientific Research Articles*, Amsterdam, John Benjamins.
- Hyland, Ken, 2000, *Disciplinary Discourses: Social Interactions in Academic Writing*, Harlow, Longman.
- Mauranen, Ann, 2004, "Where Next? A Summary of the Round Table Discussion".

- In: Tognini-Bonelli, Elena / Del Lungo Camiciotti, Gabriella (eds.), *Strategies in Academic Discourse*. Amsterdam, John Benjamins Publishing: 269-374.
- Martín-Martín, Pedro, 2005, *The Rhetoric of the Abstract in English and Spanish Scientific Discourse*, Bern, Peter Lang.
- Myers, Geoffrey, 1985, "Texts as Knowledge Claims: The Social Construction of Two Biology Articles". *Social Studies of Science* 15/4: 593-630.
- Myers, Greg, 1990, *Writing Biology: Texts in the Social Construction of Scientific Knowledge*, Madison, University of Wisconsin Press.
- Scott, Mike, 2004, *WordsmithTools 4.0*, Oxford, Oxford University Press.
- Salager-Meyer, Françoise, 1990, "Discoursal Flaws in Medical English Abstracts: A Genre Analysis per Research and Text Type". *Text* 10/ 4: 365-384.
- Sinclair, John, 1996, "The Search for Units of Meaning". *Textus* 9/1: 75-106.
- Sinclair, John, 2004, *Trust the Text: Language, Corpus and Discourse*, London, Routledge.
- Swales, John, 1990, *Genre Analysis: English in Academic and Research Settings*, Cambridge, Cambridge University Press.
- Thompson, Geoff / Ye, Yiyun, 1991, "Evaluation in the Reporting Verbs Used in Academic Papers". *Applied Linguistics* 12/4: 365-382.
- Tognini-Bonelli, Elena, 2001, *Corpus Linguistics at Work*, Amsterdam, John Benjamins.
- White, Peter R.R., 2001, *Appraisal Outline*. Available at <http://www.grammatics.com/appraisal/AppraisalOutline/AppraisalOutlineWPFiles.html> [October 15, 2008].