

Detecting key sentences for automatic assistance in peer reviewing research articles in educational sciences

Ágnes Sándor

Xerox Research Centre Europe
6. ch. Maupertuis 38240 Meylan,
France
A.Sandor@xrce.xerox.com

Angela Vorndran

DIPF
Schlossstrasse 29, 60486 Frankfurt,
Germany
vorndran@dipf.de

Abstract

The evaluation of scientific performance is gaining importance in all research disciplines. The basic process of the evaluation is peer reviewing, which is a time-consuming activity. In order to facilitate and speed up peer reviewing processes we have developed an exploratory NLP system in the field of educational sciences. The system highlights key sentences, which are supposed to reflect the most important threads of the article. The highlighted sentences offer guidance on the content-level while structural elements – the title, abstract, keywords, section headings – give an orientation about the design of the argumentation in the article. The system is implemented using a discourse analysis module called concept matching applied on top of the Xerox Incremental Parser, a rule-based dependency parser. The first results are promising and indicate the directions for the future development of the system.

1 Introduction

With the increase of centrally allocated research funding, the growing number of conferences, workshops and journals, the evaluation of scientific articles has become a central problem of the scientific community (see for example Whitley and Gläser, 2007). The evaluation of articles consists in peer reviewing, i.e. peers' reading, understanding and commenting the articles. The peer reviewing process is a matter of extensive research (e.g. Bornmann 2003, Lu 2005, 2008) discussing its reliability and evaluation methods.

Peer reviewing is a very time-consuming assignment, and Natural Language Processing (NLP) technologies might provide tools that

could shorten the time that peer reviewers take to process the articles.

Within the 7th framework EU project, European Educational Research Quality Indicators (<http://www.eerqi.eu>), we have set up this goal, and are developing a tool for providing assistance to peer reviewers in educational sciences. We do not know of any other work with this perspective.

Our approach consists in highlighting key sentences in the articles that can be regarded as the logical backbone of the article. Our tool does not evaluate, but aims at focusing the evaluator's attention on the parts of the texts that are relevant as a basis for his/her judgment. Nor does this tool check if the texts conform to some formal norms of scientific writing.

We regard highlighting key sentences as a complement to the processing guidance that the structural layout of the articles provides. The structural layout of scientific articles – title, abstract, keywords, section headings – guide the reader in processing the logical, argumentative and content-wise development of the article at different levels: The title is the brief indication of the topic, the keywords yield the conceptual context of the topic, the abstract provides a concise summary of the problems and results, and the section headings guide the reader step by step in the development of the article. Besides these waymarkers, the highlighted key sentences are meant to be an intermediary representation of content development between the title, the keywords, the abstract and the section headings on the one hand and the whole article on the other hand.

Since we define key sentences as those sentences that sum up the main messages of the articles, and since peer reviewing consists in judging the scientific value of the main messages, we

assume that highlighting key sentences both helps understanding and provides evidence for the peer reviewer's evaluation. By highlighting we intend to add a relevant and coherent dimension of the representation of the flow of the article, which is otherwise hidden, and which the reader has to discover in order to understand the article.

Highlighting is carried out using the Xerox Incremental Parser (XIP), a rule-based dependency parser (Ait-Mokhtar et al., 2002).

We will first provide a brief review of related work. This is followed by the description of the role of structural layout in educational research articles, which we wish to complement by highlighting sentences. In the subsequent sections we define the attributes of key sentences that serve as a basis for their detection and describe the natural language processing system. In the succeeding section we present our first tests for validating our approach, and finally we draw some conclusions and indicate the directions in which we plan to carry on this work.

2 Related work

Our work is in line with the growing amount of research in documentation sciences and natural language processing that takes into account the argumentative structure of research articles in tasks such as information retrieval, information extraction, navigation within documents and summarization.

In the domain of information retrieval as far back as the beginning of the 1990's Liddy (1991) claimed that additional functions for search instruments could benefit from including the discourse-level context of the retrieved search terms in the interpretation of the results. Liddy stressed the "semantic roles" of concepts in a document as opposed to the simple occurrence of search terms. Oddy et al. (1992) proceed in this line of research and state that discourse-level structures in research texts could be useful to support retrieval for the user because they represent structural qualities recognized by the reader independent of the topic of the research. Both concentrate on the analysis of abstracts of research articles and propose a system to combine topical with structural information in the retrieval process.

Kando (1997) also emphasizes the importance of the discourse-level context of search terms in the retrieved documents. The allocation of retrieved passages to functional units and thus

the possibility to gain information about article structures provides a valuable opportunity to improve the user's assessment of the retrieved documents. A similar method of annotating text passages according to their function in the text is conducted by Mizuta et al. (2006) with the objective of categorizing articles in different document genres.

Teufel and Moens (2002) base automatic summarization on extracting sentences annotated with respect to their discourse function in the text.

Lisacek et al (2005) detect sentences in biomedical articles that describe substantially new research based on analyzing discourse functions.

Another line of research to exploit the argumentative structure for navigation and information extraction is inspired by the semantic web. Instead of automatically discovering argument structures in texts, the approach aims at creating conceptually motivated processing editors in which the users insert content according to its argumentative function. (see for example Uren et al., 2007, Couto and Minel, 2007.)

3 The structure of educational research articles

Research articles in the educational sciences tend to display a very heterogeneous structure, like articles in many other fields in social sciences and humanities. While the thematic contents of the articles are structured according to the requirements of the topic, frequent occurrences of a unifying structure are introductory and concluding chapters. However, where these chapters appear they do not display uniform headings (cf. Fiedler, 1991:98). Likewise Ruiying and Allison (2004) show, for example, that the structure of research articles in linguistics is does not conform to a common model, and section headings in many cases do not refer to the function of the chapter but to the thematic contents. Brett (1994) and Holmes (1997) observe basic structural features in the articles in political sciences and sociology. They state, however, that the section headings are usually not standardized.

In contrast to the heterogeneity of the structure and section headings of research articles in social sciences and humanities those in the hard sciences show a relatively uniform structure, and often follow the well-known pattern of Introduction – Methods – Results – Discussion, which renders their reading easier.

The structural heterogeneity of social science and humanities research articles, and particularly those within educational sciences, derives from the coverage of a wide range of research problems and the consequential variation the methods applied. This discipline includes theoretically embedded discussions as well as empirical studies or material for school praxis. These differences in the referenced subjects are reflected in the way the research articles are organized and presented. Montesi and Owen (2008:151) notice a high grade of liberty granted by the educational sciences journals for the presentation of submitted papers. They also describe a clear distinction between qualitative and quantitative approaches in research articles, the latter displaying a closer connection in structural aspects to the exact sciences than the former.

In the framework of this study we compared the structural properties of fifteen articles from three journals: the British Journal of Educational Studies (BJES), the Educational Psychology Review (EPR) and the International Journal of Educational Research (IJER). These are educational research journals covering a wide variety of topics from educational psychology to school instruction. We have made the following observations:

- a) Some section headings follow the functional structuring of natural science articles, some do not. About half of the articles contain an 'Introduction' and/or a 'Conclusion', one third has a 'Methods' section and 26% of the articles has a section entitled 'Results', 'Findings' or 'Conclusion'. Thus a basis for a functionally orientated article structure can be perceived in the first and last chapters of most of the articles. Nearly 60% of the section headings, however, are oriented towards aspects of the content of the articles and show no predefined form.
- b) All of the articles are preceded by an abstract and eleven of them have keywords assigned to them.

The keywords play an important role in our highlighting approach, since they are supposed to convey the basis for topical relevance. The number of keywords assigned per article is between two and nine. While some keywords are applied only a few times in the article, others are used 60 or even over 100 times. In some cases the keywords are very common words ('teachers', 'education') and they are used frequently throughout the text. In these cases the highlighted sentences are

supposed to indicate relevant, terminological uses of those common, non-specialised words. In other cases the keywords are rare, but they are terms used in reduced contexts, for example, terminological expressions related to the field of research. Those are very useful for a quick overview over the research topic. Keywords appearing very rarely or not at all often belong to a more general level of terminology.

From an information extraction point of view the importance of the terms in the thread of the article is known to be related to their places of occurrence: in the title, the abstract, the section headings or even in the titles of the bibliography terms have more significance than in the rest of the article. This property of terms is used in search options in digital libraries. An appearance of the query term in the introduction or conclusion could also be a hint for the term being relevant for the scientific context or the results of the study whereas terms referring to the methodology or rather non-specific terms do not convey much information about the central contents of the text.

- c) The abstract is supposed to sum up the most important aspects of a research article. The articles analyzed show that in general the sentences in the abstract correspond to assertions made throughout the articles in most of the different sections. In a few cases most sentences of the abstract were also taken up in the introductory or concluding part of the article with a summarizing function.

In this section we have shown that owing to the large number of research fields in educational sciences there is a high variety in the structural design and organisation of the contents of educational science research articles. In contrast to research literature in the natural sciences, the understanding of educational sciences articles is not promoted by predefined structuring of the contents. Additionally, a terminological vagueness sometimes stands in the way of using keywords as reliable content indicators. In our approach we therefore aim at a representation of article contents independent of the structural properties of the articles.

4 The detection of key sentences

In defining the characteristic features of key sentences that serve as a basis for their detection we rely on the kinds of judgments peer review

evaluations are supposed to make (Bridges 2008).¹ We have summed up these judgments as follows: the relevance of the topic, the clarity of the problem statement, the coherence of the argumentation and the well-foundedness of the conclusions. These criteria of judgment are often presented as questions in the evaluation forms that peer reviewers are asked to fill in. Based on these evaluation criteria we define key sentences as sentences that describe research problems, purposes and conclusions related to the topic of the articles as indicated by the keywords.

The key sentences receive two types of labels in our system: SUMMARY – the sentences that convey either the goal or the conclusion - or PROBLEM – the sentences that mention research problems. Some sentences get both labels. Labeling is carried out by rules, which rely on the conceptual definition of SUMMARY and PROBLEM sentences as we show below.

In order to explain the conceptual definition we present a series of examples. The following SUMMARY and PROBLEM sentences are the first and last three key sentences detected in the same article (Barrow, 2008). In the first series of examples the keywords are underlined:

Beginning:

- (1) PROBLEM: The most challenging questions concern whether the body provides an alternative route to knowledge, if so of what.
- (2) PROBLEM_SUMMARY I do not question this belief, but in this paper I shall try to differentiate between and evaluate a number of quite distinct claims about the importance of the body in relation to schooling in general and education in particular.
- (3) PROBLEM: However, to assume, as some philosophers would, that acceptance of that premise concludes the debate on the question of education and the body, by implicitly claiming that education has nothing to do with the body per se, would be absurd.

End:

- (4) SUMMARY: Do I therefore conclude, as rationalist philosophers of education are generally supposed to conclude, that education has nothing to do with the body?

¹ In a preliminary experiment we tried to identify key sentences in an example-based way. Six scholars marked the key sentences in four articles from four domains according to the same evaluation criteria. There were hardly any overlaps. This led us to define key sentences.

- (5) PROBLEM: Second, while most of the claims made about the body and knowledge are variously opaque, suspect, or clearly wrong, it remains true that to be fully aware of or to fully understand an art form such as ballet, you need to engage in it.
- (6) PROBLEM: More generally, let us attempt to articulate more straightforward arguments for the inclusion of sports and other forms of bodily activity in the school curriculum than obscure and unconvincing claims to the effect that they are necessary, sufficient or even directly relevant to well-developed and well-rounded educational understanding.

It is apparent from these sentences that approaching the task by providing a normalized factual extraction related to the keywords as in traditional information extraction would be both very problematic - even in an intellectual (as opposed to automatic) way - and may also be useless in the case of an article whose discipline is not related to describing facts, but rather to arguing about concepts. On the other hand, the human reader clearly seizes that these sentences do describe problems, aims and conclusions related to the underlined keywords.² In the next step we define the characteristic features of SUMMARY and PROBLEM sentences as being conveyed independently of the factual propositions.

The features of the key sentences are assigned by applying the concept-matching framework described in the following series of examples. This framework had previously been successfully used in revealing argumentative functions of research articles in a text-mining application of biomedical abstracts (Lisacek et al., 2005) and in citation-type analysis (Sándor et al., 2006). (Besides processing scientific articles, concept matching has also been used in risk detection in Sándor, 2009.)

The features of key sentences are determined by the argumentative expressions in the sentences, which in some way comment on the core factual propositions. In the next series of examples we have underlined these argumentative expressions in the same set of sentences:

² At this point we do not attempt to specify the kind of relationship between the argument types and the keywords: this relationship remains simple co-occurrence.

Beginning:

- (1) PROBLEM: The most challenging questions concern whether the body provides an alternative route to knowledge, if so of what.
- (2) PROBLEM_SUMMARY I do not question this belief, but in this paper I shall try to differentiate between and evaluate a number of quite distinct claims about the importance of the body in relation to schooling in general and education in particular.
- (3) PROBLEM: However to assume, as some philosophers would, that acceptance of that premise concludes the debate on the question of education and the body, by implicitly claiming that education has nothing to do with the body per se, would be absurd.

End:

- (4) SUMMARY: Do I therefore conclude, as rationalist philosophers of education are generally supposed to conclude, that education has nothing to do with the body?
- (5) PROBLEM: Second, while most of the claims made about the body and knowledge are variously opaque, suspect, or clearly wrong, it remains true that to be fully aware of or to fully understand an art form such as ballet, you need to engage in it.
- (6) PROBLEM: More generally, let us attempt to articulate more straightforward arguments for the inclusion of sports and other forms of bodily activity in the school curriculum than obscure and unconvincing claims to the effect that they are necessary, sufficient or even directly relevant to well-developed and well-rounded educational understanding.

The detection is based on the words underlined. The system recognizes them since they belong to a database of previously compiled sets of words. The sets correspond to more or less loosely understood semantic fields that have been found to be relevant in scholarly argumentation in the previous applications of the concept-matching framework. The compilation of the lists has been entirely manual. Starting from a small number of seed words we incrementally extend the list over subsequent analyses and testing. Having worked out a first concept-matching system, its modification for a new scholarly domain takes some weeks provided that a sufficiently large corpus is available. We are carrying out experiments for automatic enrichment with the help of Wordnet, but the results have not

been satisfactory up to this point. However, since the semantic fields concerned contain a relatively well-identifiable vocabulary within the genre of scholarly writing, most of these words can be obtained from textbooks on academic writing.

In the concept-matching framework these sets of words and expressions are called constituent concepts. In previous applications nine constituent concepts have been identified for labeling argumentative sentences (Sándor, 2007). Out of these we use five here: MENTAL, IDEA, PUBLICATION, DEICTIC, CONTRAST.

In the present system we have used all the words that have been compiled for labeling argumentative functions of biomedical research abstracts, and we have added a few others after having studied some educational research articles. Augmenting the list of words in the constituent concepts undoubtedly increases the coverage of the system, although we have found that the words already compiled yield fairly large coverage.

In terms of the constituent concepts we define PROBLEMs as CONTRASTed IDEAs or CONTRASTs in MENTAL operations involved in research, while SUMMARIES of one's research goals and conclusions consist in pointing out in the current (DEICTIC) PUBLICATION one's (DEICTIC) IDEAs or MENTAL operations. We cite now the example sentences only through the constituent concepts of PROBLEM and SUMMARY:

Beginning:

- (1) PROBLEM: ... challenging[C,M] questions[C,M] ... whether[C] ... alternative[C] ... to knowledge[I] ...
- (2) PROBLEM_SUMMARY: I[D] ... question[C,M] this belief[M] ... in this[D] paper[P]...
- (3) PROBLEM: However[C] to assume[C,M], ... that acceptance[MC] ... concludes[C,M] the debate[C,I] ..., by ... claiming[C,M] ... would be absurd[C].

End:

- (4) SUMMARY: ...I[D] ... conclude[C,M] ...
- (5) PROBLEM: ... while[C] ... the claims[I] ... are ... wrong[C] ...
- (6) PROBLEM: ... unconvincing[C,M] claims[I] ...

It is apparent that the words that represent the constituent concepts in these sentential skeletons constitute purely argumentative expressions and are void of any factual proposition.

However, not all sentences containing these words convey the target concepts. Consider for example the following sentence from a research article (Meinberg and Stern, 2003.):

- (7) Only 1.8% of the claims were attributed to wrong-site surgery, but 84% of the claims due to wrong-site surgery resulted in payment to the plaintiff compared ...

In order to differentiate between relevant and irrelevant ways of combining the constituent concepts in a sentence our framework proposes syntactic criteria: sentences are labeled in case the constituent concepts are in syntactic dependency relationship with each other. The kind of syntactic relationship is not specified.

The restriction of syntactic dependencies is especially relevant in the case when the constituent concepts are function words (like e.g. *not*) or have a general sense (like e.g. *work*). At this point we have not measured the impact of this restriction on recall and precision.

We have built the concept-matching grammar for labeling argumentation types on top of a general-purpose dependency grammar developed in XIP. In the concept-matching grammar we define the argumentative expressions as those syntactic dependencies where both words belong to the particular concepts that constitute the target concepts as defined above. The only exceptions to the syntactic constraint are sentential adverbs (like “however”), for which the XIP grammar does not extract any syntactic dependencies. The highlighted sentences are those that contain the labeled argumentative expressions.

5 First tests

Our exploratory system is based on several consecutive hypotheses, the validity of which should be tested incrementally.

The first hypothesis is that the key sentences relevant for peer reviewing are those that describe the problems, aims and results in the articles, and that these sentences contain the keywords provided with the articles. The second hypothesis is that these sentences can be detected using the concept-matching grammar. Finally the third hypothesis is that highlighting these sentences can save peer reviewers’ time evaluating articles.

Owing to the complexity and relative vagueness of the task, we have not been able to set up either a formal or a statistically significant evaluation up to now. For this article we have

carried out an initial internal test³ towards the validation of the first two hypotheses.

In a test corpus of five articles from the three educational research journals mentioned in Section 3 (BJES, EPR, IJES) we checked if the sentences highlighted by the system convey relevant information in the argumentative development of the paper and if we find other key sentences that are not highlighted. Next we analyzed the causes of silence and noise in order to evaluate our basic assumptions.

Table 1 summarizes the results of the test over the five articles in terms of recall and precision of the key sentences, and also indicates the percentage of key sentences out of all the sentences in the articles. Recall is defined as the number of correct sentences highlighted divided by the total number of sentences that we considered to be key sentences. Precision is defined as the number of correct sentences highlighted divided by the total number of sentences highlighted.

Table 2 shows if the missing sentences identified as key sentences by the evaluator contain keywords or not. It also displays the number of missing sentences in each article by type of error. Table 3 shows the number of false positive sentences according to the types of the causes of the error.

Article	Recall	Precision	Key sentences (Number of sentences)
BJES-1	77%	67%	17% (195)
BJES-2	69%	77%	11% (240)
EPR-1	39%	59%	8% (331)
EPR-2	30%	100%	3% (330)
IJER-1	35%	67%	2% (526)

Table 1. Recall and precision of key sentences detected and percentage of key sentences out of all the sentences in the article

³ This test was carried out by one of the co-authors of this article who did not take part in the development of the NLP system.

Article	Keywords in sentence		1	2	3
	yes	no			
BJES-1	6	1	4	2	3
BJES-2	5	1	1	-	5
EPR-1	13	12	3	6	16
EPR-2	13	10	-	-	23
IJER-1	8	3	-	1	10
All	45	27	8	9	57

Table 2. Causes of silence: 1.Incorrect analysis by the parser; 2.Inadequacy of the framework for the task; 3. Not SUMMARY or PROBLEM sentence according to our definition

Article	1	2	3
BJES-1	6	1	3
BJES-2	1	2	4
EPR-1	6	-	5
EPR-2	-	-	-
IJER-1	1	-	2
All	14	3	14

Table 3. Causes of noise: 1.The sentence matches the rules but is not important enough; 2.Incorrect analysis by the parser 3.Inadequacy of the framework for the task

We can observe significant differences according to the journals with respect to both hypotheses that we have tested. The three journals deal with rather different research topics ranging from theoretical discussions to empirical studies of students' behavior. According to our results the important passages of these articles are characterized by different attributes: while in empirical studies more or less definite results can be presented, theoretical discussions rest more on a discursive level offering less clear conclusions to be identified as SUMMARY or PROBLEM sentences. This is reflected on the one hand in the differences in recall and precision among the journals and on the other hand in the differences in the number of sentences missing due to error-type 3 in Table 2.

In the EPR and in the IJER we found more key sentences that are neither SUMMARY nor PROBLEM sentences according to our definition than in the BJES. Most of these sentences convey definitions related to the key concept. Thus our first hypothesis seems to hold more for empirical studies than for theoretical ones. In order to increase the coverage of key sentences the

system should be completed so that it also detects definitions, especially in the case of theoretical articles.

As for the presence of keywords in the key sentences, our results show that this is a relevant condition, however not necessary since a number of key sentences identified do not contain keywords. Further study is needed to identify the characteristic features of key sentences without keywords. We have carried out an additional test to see if the correct key sentences cover all the keywords in the list. In the five articles we have only found one keyword that was not present in any of the key sentences, but this word appeared only once in the whole article. The fact that relatively few sentences are detected in the articles and that in these sentences all the keywords are covered supports the hypothesis that the key sentences do play an important role in the thread of the article.

Among the errors leading to both silence and noise we have found a number of cases where the concept-matching framework in its present form is not convenient for the task of detecting key sentences that satisfy the conditions or filtering erroneous sentences. The reason for this in both cases is that the unit of concept-matching is the sentence, whereas in these cases a single sentence does not provide enough context for identifying or for specifying the target concepts respectively. Since the number of errors due to this reason is not very high we do not consider that these results invalidate the second hypothesis. The number of such false positives is quite significant, however, which might be disturbing for the user of the system. This kind of error could be overcome by enlarging the scope of concept-matching beyond the sentence. In this way we could filter out these false positives.

In a significant number of cases noise is not due to an error in the system but to the fact that the sentence is not important enough with respect to the development of the whole article. Whether this kind of noise is a significant disturbing factor for the user is to be tested in subsequent evaluation by users.

Finally, we have found few errors due to bugs in the grammar, which indicates that the recognition of SUMMARY and PROBLEM sentences is relatively reliable. These results also contribute to suggesting that the second hypothesis seems to hold.

6 Conclusion

In this article we have presented an exploratory system for highlighting key sentences containing keywords in order to support peer review. The selected sentences are supposed to help peer reviewers of articles in educational sciences to focus their attention on some relevant textual evidence for formulating their judgments. We have argued that even if the structural characteristics—the abstract, the keywords and the section headings—guide the reader in following the development of the article, content-oriented highlighting of key sentences might enhance the quick understanding of the core contents.

Although the subjects of educational science research articles display very heterogeneous structures and contents, the system could identify a number of sentences containing the main statements of the articles. Text-inherent developments not accompanied by structural signs like the outcomes of empirical studies or the contents of a theoretical discussion about abstract terms could be identified using automatic text analysis, which can possibly save intellectual effort of scientists. The time-consuming task of reviewing a growing number of research publications, hardly manageable when studying each submitted manuscript thoroughly, could thus be facilitated and supported and less threatened to be replaced by wholly automatic metric systems when time constraints become more severe.

The method we have developed is implemented in XIP, a rule-based dependency parser. It uses pre-existing lexical resources and applies the concept-matching framework.

The results of our first tests suggest that two of our three initial hypotheses are partially valid. According to our first hypothesis the key sentences relevant for peer reviewing are those that describe the problems, aims and results in the articles. We have found that sentences conveying definitions, especially in theoretical articles, should also be highlighted as key sentences. The second hypothesis is that these sentences can be detected using the concept-matching grammar. We have found in the majority of cases that this hypothesis is valid, however, enlarging the unit of concept-matching to multiple sentences would improve the performance.

Based on this result we are undertaking a user evaluation to measure the time needed to peer review these articles with and without highlighting. We are also planning to extend the system

in the two directions suggested by the test results.

Besides providing assistance to peer reviewers the system presented here could be used in other applications, which we would like to explore in future projects. The possibilities include improving search functionalities in digital libraries, displaying electronic documents by linking keywords to key sentences and discourse-based navigation.

Acknowledgments

The research leading to these results has received funding from the European Community's Seventh Framework Programme under grant agreement n° 217549.

We would like to thank Alexander Botte, Aaron Kaplan, Peter Meyer and our partners in the EERQI project for their valuable contributions and suggestions.

References

- Salah Ait-Mokhtar, Jean-Pierre Chanod, and Claude Roux. 2002. Robustness beyond shallowness: incremental dependency parsing. *Natural Language Engineering*, 8(2/3):121-144.
- Robin Barrow. 2008. Education and the Body: Prolegomena. *British Journal of Educational Studies* 56(3):272-285.
- Lutz Bornmann and Hans-Dieter Daniel. 2003. Begutachtung durch Fachkollegen in der Wissenschaft. Stand der Forschung zur Reliabilität, Fairness und Validität des Peer-Review-Verfahrens. *Universität auf dem Prüfstand. Konzepte und Befunde der Hochschulforschung*. (S. Schwarz and U. Teichler, Eds.). Campus Verlag Frankfurt/New York: 207-225.
- Paul Brett. 1994. A genre analysis of the results section of sociology articles. *English for Specific Purposes*, 13(1):47-59.
- David Bridges. 2008. Criteria of Quality in Educational Research. Working Group Report of the 1st EERQI Workshop, 20-21 June 2008. Leuven. Project Internal Document.
- Javier Couto and Jean-Luc Minel. 2007. *NaviTexte : a Text Navigation Tool*. Artificial Intelligence and Human-Oriented Computing, Lecture Notes in Artificial Intelligence, 4733, Springer, Berlin, Heidelberg.
- Susanne Fiedler. 1991. *Fachtextlinguistische Untersuchungen zum Kommunikationsbereich der Pädagogik dargestellt an relevanten Fachtextsorten im Englischen*. Lang, Frankfurt a.M.

- Richard Holmes. 1997. Genre analysis, and the social sciences: An investigation of the structure of research article discussion sections in three disciplines. *English for Specific Purposes*, 16(4):321-337.
- Noriko Kando. 1997. Text-level structure of research papers: Implications for text-based information processing systems. *Proceedings of the 19th British Computer Society Annual Colloquium of Information Retrieval Research*, Sheffield University, Sheffield, UK, 68-81.
- Elizabeth D. Liddy. 1991. The discourse-level structure of empirical abstracts: an exploratory study. *Information Processing and Management*, 27(1):55-81.
- Frédérique Lisacek, Christine Chichester, Aaron Kaplan, and Ágnes Sandor. 2005. Discovering paradigm shift patterns in biomedical abstracts: application to neurodegenerative diseases. *First International Symposium on Semantic Mining in Biomedicine*, Cambridge, UK, April 11-13, 2005.
- Yanping Lu. 2005. Editorial Peer Review in Education: Mapping the Field. *Australian Association for Research in Education 2004 conference papers, Melbourne, Australia (Jeffery, P. L., Ed.):*1-19.
- Yanping Lu. 2008. Peer review and its contribution to manuscript quality: an Australian perspective. *Learned Publishing*, 21(3):307-316.
- Eric G. Meinberg and Peter J. Stern. 2003. Incidence of Wrong-Site Surgery Among Hand Surgeons. *The Journal of Bone and Joint Surgery (American)* 85:193-197.
- Yoko Mizuta, Anna Korhonen, Tony Mullen, and Nigel Collier. 2006. Zone analysis in biology articles as a basis for information extraction. *International Journal of Medical Informatics*, 75(6):468-87.
- Michaela Montesi and John Mackenzie Owen. 2008. Research journal articles as document genres: exploring their role in knowledge organization. *Journal of Documentation*, 64(1):143-167.
- Robert N. Oddy, Elizabeth D. Liddy, Bhaskaran Balakrishnan, Ann Bishop, Joseph Elewononi and Eileen Martin. 1992. Towards the use of situational information in information retrieval. *Journal of Documentation*, 48(2):123-171.
- Yang Ruiying and Desmond Allison. 2004. Research articles in applied linguistics: structures from a functional perspective. *English for Specific Purposes*, 23(3):264-279.
- Ágnes Sándor, Aaron Kaplan and Gilbert Rondeau. 2006. Discourse and citation analysis with concept-matching. *International Symposium: Discourse and document (ISDD)*, Caen, France, June 15-16, 2006.
- Ágnes Sándor. 2007. Modeling metadiscourse conveying the author's rhetorical strategy in biomedical research abstracts. *Revue Française de Linguistique Appliquée* 200(2):97-109.
- Ágnes Sándor. 2009. Automatic detection of discourse indicating emerging risk. To appear in *Critical Approaches to Discourse Analysis across Disciplines. Risk as Discourse – Discourse as Risk: Interdisciplinary perspectives*.
- Simone Teufel and Marc Moens. 2002. Summarizing scientific articles: experiments with relevance and rhetorical status. *Computational Linguistics*, 28(4):409-445.
- Victoria Uren, Simon Buckingham Shum, Clara Mancini, and Gangmin Li. 2007. Modelling Naturalistic Argumentation in Research Literatures: Representation and Interaction Design Issues. *International Journal of Intelligent Systems, (Special Issue on Computational Models of Natural Argument, Eds: C. Reed and F. Grasso)*, 22(1):17-47.
- Richard Whitley and Jochen Gläser. 2007. *The Changing Governance of Sciences: The Advent Of Research Evaluation Systems*. Springer