A Multi-Layered Annotated Corpus of Scientific Papers

Beatriz Fisas, Francesco Ronzano, Horacio Saggion

DTIC - TALN Research Group, Pompeu Fabra University

c/Tanger 122, 08018 Barcelona, Spain

{beatriz.fisas, francesco.ronzano, horacio.saggion}@upf.edu

Abstract

Scientific literature records the research process with a standardized structure and provides the clues to track the progress in a scientific field. Understanding its internal structure and content is of paramount importance for natural language processing (NLP) technologies. To meet this requirement, we have developed a multi-layered annotated corpus of scientific papers in the domain of Computer Graphics. Sentences are annotated with respect to their role in the argumentative structure of the discourse. The purpose of each citation is specified. Special features of the scientific discourse such as advantages and disadvantages are identified. In addition, a grade is allocated to each sentence according to its relevance for being included in a summary.To the best of our knowledge, this complex, multi-layered collection of annotations and metadata characterizing a set of research papers had never been grouped together before in one corpus and therefore constitutes a newer, richer resource with respect to those currently available in the field.

Keywords: multi-layered annotated corpus, scientific discourse, citations, summarization gold standard

1. Introduction

The development of natural language processing tools for information extraction or document summarization tailored to scientific literature will provide quick tracking of scientific creativity and innovation. Easy access to challenges faced by the researchers, their results and contributions, and how these relate to the works of other researchers highlighting the novelties or advantages of the explored scientific project may inspire new approaches in a line of investigation.

With the aim of supporting automated analysis and thus easier access to this information, we have generated a multi-layered annotated corpus of scientific discourse. In this article, after introducing our multi-layered scienfic annotation schema, we describe the way we collaboratively annotate our corpus so as to create its gold standard version.

Corpus annotations have been provided in two stages. In the first stage (Fisas et al., 2015), annotators were asked to characterize the argumentative structure of papers by associating each sentence to one over 5 categories (Challenge, Background, Approach, Outcome and Future Work) and eventually specifying for Challenge sentences a subcategory among Goal and Hypothesis and distingushing among the Outcome sentences the ones that describe an authors' Contribution. Based on the work of Liakata et al. (2010) and Teufel (2010) we developed an annotation schema and produced an annotated corpus. Its quality was evaluated in terms of the inter-annotator agreement (K=0.66) comparable to the values attained by the aforementioned researchers.

The results were analysed by category and 5 main areas were identified in the articles where the middle 50% of sentences included in each category were located. The output was a gold standard annotated corpus (10,777 sentences; 40 documents) in the domain of Computer Graphics. This corpus constitutes a valid dataset to experiment with automatic sentence classification algorithms.

In the second phase of the annotation, a step further into some other aspects of the scientific discourse has been taken. In the first place, we detected the interplay between the author's work and other researchers' contributions in the field by annotating the purpose of citations. In the second place, we tried to identify some frequent features in the scientific discourse (Advantages, Disadvantages, Novelties, Common practices and Limitations), which will allow to improve the comparison of articles in the domain. Simultaneously, we graded sentences in terms of their relevance for a summary, in order to provide a manually annotated resource for reference and training an automatic summarization tool.

2. State of the Art

In this Section we provide a brief overview of the most relevant annotation schemas proposed to support the characterization of citations, the identification of relevant scientific discourse statements and the spotting of text excerpts useful to summarize a paper.

Dealing with citation characterization, Moravcsik and Murugesan (1975), precursors in this research field, presented an in depth study of the nature of citations and developed a typology to estimate their quality and context. The work of Jörg (2008), following Moravcsik and Murugesan (1975), inspired the Citation Typing Ontology (CiTO), who was later evaluated in terms of its use for annotation by Ciancarini et al. (2014).

The other most influential taxonomy is proposed by Spiegel-Rösing (1977), with 13 categories. However, 80% of the citation purposes could be classified in one category: *Cited source substantiates a statement of assumption, or points to further information*.

Nanba and Okumura's contribution (1999) is a very simplified scheme with 3 categories (Basis, Comparison and Other).

Teufel et al. (2006), who consider citations as signals of knowledge claims in the discourse structure, introduced a citation annotation scheme, with 12 categories, adapted from Spiegel-Rösing (1977), and inspired by the findings from Swales (1990) that scientific argument follows a general rhetorical structure to study the interplay of discourse structures of scientific arguments with formal citations.

Following Spiegel-Rösing (1977) and Teufel et al. (2006), Abu-Jbara and Radev (2012) stay with 6 categories (Criticism, Comparison, Use, Substantiation, Basis and Neutral) to determine the purpose and polarity of citations.

Research papers include the description of concepts such as advantages, disadvantages or novelties that do not belong exclusively to any of the structural sections of the discourse. They are useful for comparison between scientific articles. Both Liakata et al. (2010) and Teufel et al. (2009) have incorporated some crosswise features in their annotation schemes. Liakata's 3-layered annotation scheme devotes the 2nd layer to the annotation of properties of some of the concepts previously identified in the first layer. AZ-II, Teufel's annotation scheme, defines a category to characterize a novelty or an advantage of the approach mentioned in the paper.

In reference to grading sentences for summarization, Saggion et al. (2002) compiled human-generated "ideal" summaries at different compression rates, and obtained a gold-standard of sentence-based agreement, both between the annotators, and between the summarizer and the human annotators. Sentences were assigned a score from 0 (irrelevant) to 10 (essential) expressing the annotators subjective opinion about how relevant each sentence is for a summary.

3. Multi-Layered Annotation Schema

3.1. Citations: Purposes and Subpurposes

Our annotation scheme for citation purposes is an extension of the proposal of Abu-Jbara and Radev (2012), a well-balanced selection of 6 top-categories, to which a second level of sub-purposes has been added (Table 1).

The sub-purposes' motivations are diverse: *Weakness* and *Strength* include a polarity judgement, *Evaluation* intends to collect those sentences where a balance of a positive and a negative comment on a cited paper is expressed; *Similarity* and *Difference* are opposite reasons for comparison. The sub-purposes suggested for the purpose **Use**, are different elements of a cited work that can be used by the author of the citing work (see Table 1). Citations categorised as **Basis** include the reference to the works of researchers upon which the citing work builds but also to the author's *Own work*; some cited works may also be suggested for *Future work*. Finally, the **Neutral** category includes all the other citations that can be a mere *Description* of a researcher's work, the *Reference for more information* or even a comment on *Common practices* in the field.
For example, the citation in the sentence:

PURPOSE	SUB-PURPOSES
CRITICISM	Weakness
	Strength
	Evaluation
	Other
COMPARISON	Similarity
	Difference
USE	Method
	Data
	Tool
	Other
SUBSTANTIATION	
BASIS	Previous own Work
	Others work
	Future Work
NEUTRAL	Description
	Ref. for more information
	Common Practices
	Other

Table 1: Citation Purpose Annotation Scheme

Our approach is similar to margin-based linear structures classification [Taskar et al, 2003]

is classified as purpose: **Comparison** and subpurpose: *Similarity*.

3.2. Crosswise Features

Based on the previous work of Teufel and Liakata, our annotation aim is to detect characteristic features of the scientific discourse that may appear at any point in a research paper. Therefore, the annotation scheme includes the following 7 categories: **Advantage** and **Disadvantage** not only limited to the author's approach but also to any reference to an advantage or disadvantage in the documents of our corpus. Since advantages and disadvantages appear frequently in the same sentence, we have also included two double categories: **Advantage-disadvantage** and **Disadvantage-advantage**. Scientific literature also pays special attention to **Novelties** (not exclusive of the author's approach) and comments on **Common practices** in the field, so these concepts were included in the annotation scheme. Finally, **Limitations** (only referred to the author's work) are also tagged, as they are of paramount importance in the comparison of different investigations.
For example, the sentence:

Skeleton Subspace Deformation (SSD) is the predominant approach to character skinning at present.

is classified as containing a **Common Practice.**

3.3. Grading for summarization

The third annotation task is related to the summarization of scientific documents, following the works of Saggion et al. (2002) and Radev et al. (2003).

GRADE	DEFINITION
1	TOTALLY IRRELEVANT FOR A SUMMARY
2	SHOULD NOT APPEAR IN A SUMMARY
3	MAY APPEAR IN A SUMMARY
4	RELEVANT FOR A SUMMARY
5	VERY RELEVANT FOR A SUMMARY

Table 2: Grading Scale

The annotatorion includes a double task: grading the sentences in each document according to their relevance for being included in a summary and providing a handwritten summary no longer than 250 words.

We adopted a shorter sentence relevance grading scale than Radev et al. (2003) and asked the annotators to mark the sentences with a value from 1 to 5, 1 being the lowest relevance value and 5 the highest relevance value (Table 2).

4. Corpus Dataset and Annotation Process

As described in Fisas et al. (2015), the corpus is a set of 40 randomly selected articles among a representative sample of research papers previously chosen by experts in the domain of Computer Graphics. Articles were classified into four important subjects in this area: Skinning, Motion Capture, Fluid Simulation and Cloth Simulation.

The annotation is sentence based as we have considered sentences to be the most meaningful minimal unit for the analysis of scientific discourse.

The annotation process is characterized by its collaborative approach between the developers of the methodology, experts in annotation and text mining, and the 12 annotators who are experts in the domain of Computer Graphics. Thus, a web-based collaborative annotation tool (**Annote**) was developed, enabling users to easily annotate textual contents by exploiting a web browser (see Section 6.).

The documents were divided into 4 groups of 10 documents each, one for each of the 4 subjects included in the Computer Graphics Corpus. Each group of documents had to be annotated simultaneously by 3 annotators. Some documents were allocated for inter-annotator checking purposes.

The annotation process went through several steps:
In the first place, a training session was held with the leader annotators for each one of the 4 groups. In this training session, the designers explained the annotation goals, the motivations, tasks, categories, criteria and examples, as well as the details referred to the annotation tool and the steps to follow. The annotators were encouraged to test the tool and guidelines with a hands-on annotation workshop.

The leader annotators were then assigned a demo environment for training new annotators, together with guidelines and recommendations. Similarly, once selected, the new annotators had also a set of documents just for testing and practicing with the **Annote** Web annotation platform.

In order to monitor the progress of the annotation and detect possible deviations or difficulties, an early check was scheduled once all annotators had tagged 4 documents each.

The citation purpose annotation schema was then simplified to a coarser-grained approach as the analysis of the first results revealed that some annotators found it hard to distinguish between some sub-purposes. The schema was therefore reduced to the purposes, leaving the specification of the subpurpose as optional. At the same time, new recommendations and modificactions to the guidelines were forwarded to the annotators, making clearer priorities between some categories (for example, **Advantage** and **Disadvantage** is preferred to **Common Practice**, in the Crosswise Features task).

The last step of the process was the reconciliation of the annotated versions of the documents in order to obtain a gold standard corpus and the collection of the human summaries.

5. Annotation Guidelines and Recommendations

The annotators were provided not only with Guidelines for the annotation of the three tasks, but also with a recommended procedure for annotation.

The Guidelines provide support in the identification of the purpose of a citation, the detection of crosswise features, and the criteria for grading sentences according to their relevance for a summary. This is a tedious and hard task and requires a careful reading of the original article, and annotators are suggested to highlight the main ideas as they read on the article's hardcopy of digital copy in order to ease the grading task. Tables, figures, formulas and the division of the article in sections are dropped in the annotation's tool view of the paper.

The annotation procedure should ideally start by grading each sentence in an article, and simultaneously look for the description of an advantage, disadvantage, novelty, common practice or limitation. All sentences have a default value, which tags them as Totally irrelevant for a summary and as having no crosswise feature (label:Unselected); this default value will remain unless the annotator chooses to change it.

After the grading task, the annotators were encouraged to write their personal summary, whose length should ideally be between 200-250 words for an average article (8-10 pages). The resulting text should be a short summary of the paper.
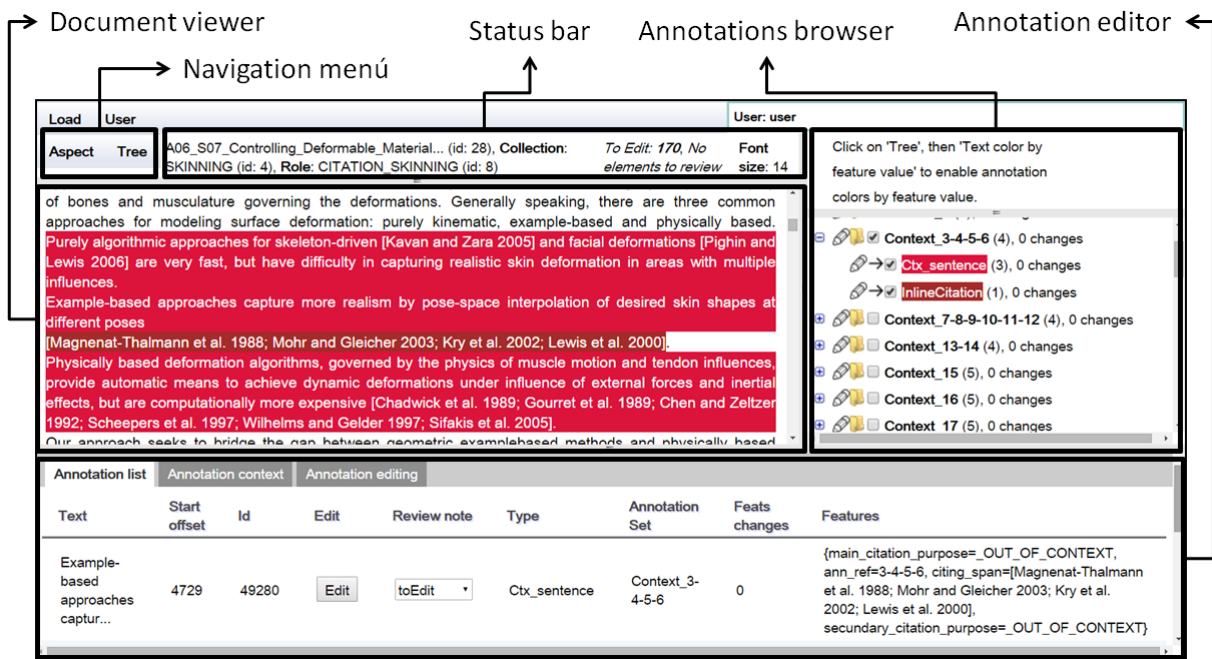
3083

Figure 1: ANNOTE: annotation Web interface.

The citation context and in-line citation are preselected by the tool in a separate collection. For each in-line citation, we considered the sentence where the citation occurs and the two sentences preceeding and following this sentence as candidate sentences for containing the purpose. The citation purpose annotation consists in reading through the whole context looking for the purpose of the citation and selecting the reason from a pop-up list.

After the early check, emphasis was made to the annotators for keeping their level of attention high and not missing information. Some modifications to the guidelines were forwarded to the whole team making priorities between categories clearer for the Crosswise Features and Citations annotation task.

For example, in the Crosswise Features task, annotators were reminded that what the author states is prior to the annotator's opinion, and that some categories are preferred to others.

Regarding the Citations task, the priorities were set such that: if a sentence can be tagged as **Criticism** (if it states an evaluation or a strength or a weakness according to the author) the annotator should prefer this category to any other. If **Use** is possible, then it will be preferred to **Basis**. Lastly **Neutral** has no preference, a citation will only be tagged as Neutral if it can't be tagged as any other category.

6. Web-based Annotation Tool: Annote

In order to enable the annotation of the our corpus by involving several annotators distributed across distinct

places, we developed **Annote**[1], a web-based collaborative annotation tool. Even if easily adaptable to carry out distinct types of textual annotation tasks, **Annote** has been customized to support the annotations of our corpus papers with respect to the facets described in details in this paper.

Textual annotations constitute the core item that annotators can create and characterize by **Annote**. Each annotation identifies a consecutive excerpt of a textual document and is characterized by its name and a set of features, like the rhetorical class or the summary relevance of the sentence identified by the excerpt. The names of the annotations as well as the named features that can characterize each annotation are specific to the annotation task. Annotations can be logically organized by grouping them into annotation sets. Usually all the annotations of an annotation set share some features (i.e. all the sentences of a section of the document).

By providing his credentials, each annotator can access **Annote** and browse a customized view of the documents he has to annotate. Annotators can annotate documents from one or more collections by choosing a specific annotation role that defines which editing actions the annotator can perform. Once selected the document to edit and the annotation role among the set of document/role pairs that are available, **Annote** document annotation view is shown to the annotator (see Figure 1).

In the center-left part of the document annotation view there is the *Document Viewer* that shows the textual contents of the document that is being annotated.

On the right side there is the *Annotation Browser* that provides a two-levels tree view of all the document annotations. The root nodes of the tree view represent the annotation sets, while each leaf identifies a name of the annotations that are contained inside the corresponding annotation set. Each annotation name is characterized by a color; when the checkbox next to an annotation name is selected, the excerpts of the text that are identified by annotations with that name are highlighted by the same color in the *Document Viewer* (see the annotations named Sentence in Figure 1).

On the top of the document annotation view there is the *Document Menu* that is useful to set properly the document visualization options and the *Document Status Bar* that shows the current annotation status of the document.

On the bottom of the document annotation view there is the *Annotation Editor*, where annotators can edit the features of the annotations and monitor the annotation status of the whole document.

Besides the general layout of the document annotation view, Figure 1 shows how Annote has been used in a specific annotation task of the Dr. Inventor Corpus: the characterization of the purpose of an in-line citation of a paper.

On the right side, in the *Annotation Browser* we can see that the the annotator is dealing with the in-line citation *[Magnenat-Thalmann et al., 1988...]* that is highlighted in the *Document Viewer* together with the sentences belonging to its context (*Ctx_sentence* annotations).

In the lower part of the document annotation view, the *Annotation Editor* shows the highlighted annotations (the sentences that belongs to the citation context) by means of a scroll list. Only the first item of this list is visualized: this item is related to the citation context sentence 'Example based approaches...'. By clicking the 'Edit' button it is possible to access the Annotation editing tab of the annotation (citation context sentence) so as to specify its citation purpose.

Annote implements several features and indicators that are useful to help annotators to keep track of both the items of the document that should be annotated and the annotations features that should be specified. In this way each annotator can monitor the progress of his annotation tasks.

7. Annotation Results

The Corpus includes 10,780 sentences, with an average of 269.7 sentences per document.

The Gold standard version has been built with totally and partially agreed sentences. The strategy adopted with the totally disagreed sentences is different in the two stages of the annotation process. In the first stage, where the
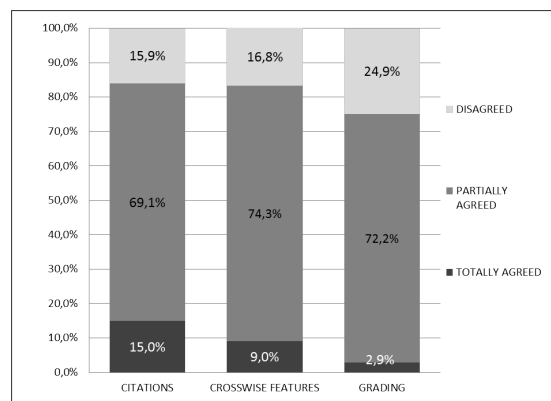

Figure 2: %Agreement in the 3 annotation tasks

sentences were classified according to their argumentative content, and only 3 annotators were involved, disagreed sentences (only 3%) were included in the Gold standard with the category chosen by the designer of the annotation scheme (who was also an annotator).

In the second stage, described in this paper, sentences with total inter-annotator disagreement have not been included in the Gold Standard because there was no reliable reference on which to base a selection among the 3 proposed categories.

The dataset also contains a triple collection of human summaries for each of the 40 documents.

7.1. Citations Gold Standard

The Gold Standard version includes the Totally and Partially agreed sentences (84%) (Fig. 2) and the distribution of the non-default categories is the following: **Criticism** 23%, **Comparison** 9%, **Use** 11%, **Substantiation** 1%, **Basis** 5% and **Neutral** 53% (Fig. 3).
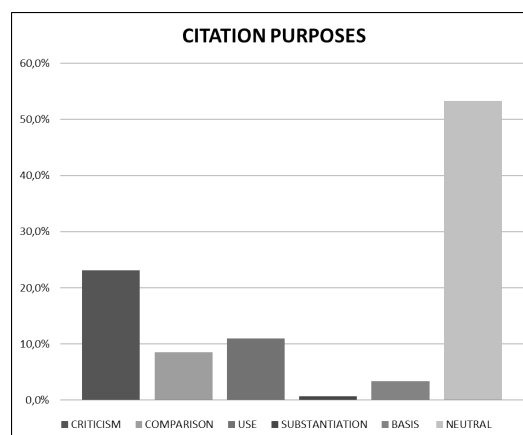

Figure 3: Distribution of purposes for citations

This distribution is comparable to the data of Abu-Jbara and Radev (2012) in a similar task of citation annotation with 30 papers.
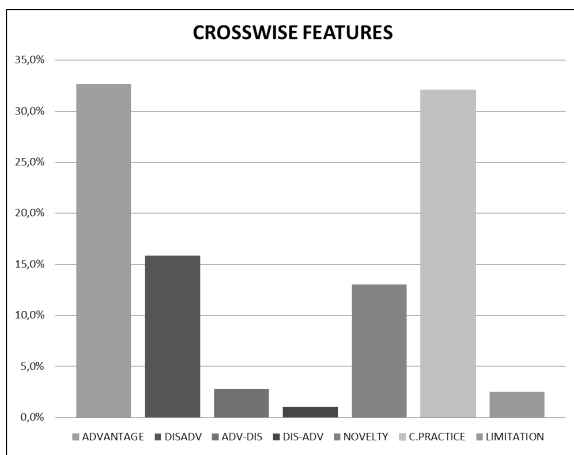
7.2. Crosswise Features Gold Standard

Figure 4: Crosswise features in the Gold Std corpus

The Crosswise Features Gold Standard contains 83% of the total sentences (Fig. 2), and excludes the totally disagreed sentences. The distribution of non-default categories in the Gold Standard is: **Advantage** 33%, **Disadvantage** 16%, **Advantage-disadavantage** 3%, **Disadvantage-advantage** 1%, **Novelty** 13%, **Common Practice** 32% and **Limitation** 2% (Fig. 4).

7.3. Grading Gold Standard

In this task, the percentage of disagreed sentences is higher (25%) (Fig. 2); however, in this case, the grade of each sentence is not a categorical feature like in previous annotations, but an ordinal one.
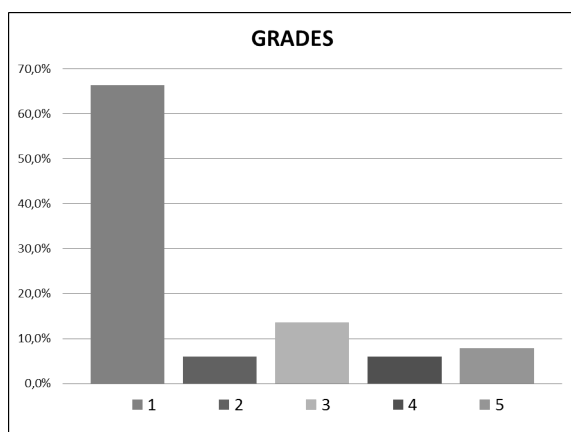

Figure 5: Distribution of sentences according to their grade

The distribution of the selected categories in the gold standard corpus of graded sentences is the following: **1-Totally irrelevant for a summary** 66%, **2-Should not appear in a summary** 6%, **3-May appear in a summary** 14% , **4-Relevant for a summary** 6% and **5-Very relevant for a summary** 8% (Fig. 5).

As expected, most sentences are considered totally irrelevant for a summary (grade **1**). However, a closer analysis of the graded sentences in one of the groups of 3 annotators (An_1, An_2, An_3) revealed 3 different styles of selecting the relevant information to be included in a summary confirming that there is not one single way of summarizing a text (Fig. 6).
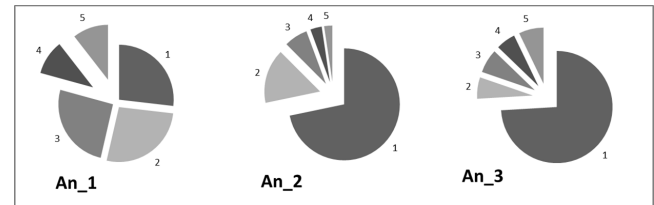

Figure 6: Average percentage of graded sentences in all the documents annotated by 3 annotators in the same team

An_2 and An_3 left the default value (grade **1**) in nearly 75% of the sentences. On the contrary, An_1 splits this 75% into sentences with grade **1**, **2** and **3**, leaving the remaining 25% equally distributed into sentences of grades **4** and **5**. An_2 considered that more than half of the graded sentences should not appear in the summary (grade **2**), a quarter could be included in the summary (grade **3**), and only the remaining quarter were considered relevant (grade **4**) or very relevant (grade **5**) for a summary in similar proportion. An_3 distributes the graded sentences in a more homogeneous way: he considered that about half of them should appear in the summary (grades **4** and **5**), while the other half corresponds to sentences of grade **2** and **3**. Finally, An_1 splitted the graded sentences into 3 thirds: according to him, one third of all the sentences should not appear in the summary (grade **2**), another third could optionally be included (grade **3**) and the last third contains the relevant sentences (one sixth) and the very relevant sentences (grade **5**).

8. Research Limitations and Future Work

The collaborative approach we have chosen was a challenge as the process implies an important amount of information being transferred and each annotator had to undergo a complex training process. In the first place, the task is multiple and diverse (annotating citations and discursive features, grading sentences, and summarizing) and, in the second place, the flux of information from the first training session, which the leader-annotators attended, had to be properly transmitted to each one of the rest of recruited annotators.

The Gold standard version of our scientific discourse corpus has been built according to the criteria of total and partial agreement among the annotators' versions. Nevertheless, the values of the inter-annotator agreement intra-group (considering only the 3 annotators of the same team) were very low for some annotators, especially in the Skinning annotation team. Further analysis must be done

in order to detect those annotators that do not meet the standard quality in the Citations and Crosswise Features annotation tasks. In this respect, two different strategies are possible: benchmarking the quality across the 4 teams against a reference annotation, or, alternatively, the most reliable annotators could be determined with MACE - Multi-Annotator Competence Estimation (Hovy et al., 2013).

The grading annotation results can be evaluated considering groups of grades, in order to detect if human annotators have difficulties in distinguishing relevant (grade **4**) and very relevant (grade **5**) sentences for a summary, or useful (grade **3**) from unnecessary information (grade **2**).

Finally, a second round in the annotation in order to re-annotate the sub-purposes which were left out after the early annotation check would now be an easier task, as the annotators are already trained, and the best ones could be selected. The annotation of the sub-purposes of citations would provide a richer resource for classifying the information contained in the citations of a scientific paper, and therefore allow a better comparison of different research activities.

9. Conclusion

We have described the motivation of our work as well as the process of design and annotation that has lead to our linguistically annotated corpus of Computer Graphics research articles, an informatively rich resource for text-mining, summarization and other NLP technologies.

As shown in Fig.7, all sentences in the corpus are classified into a Rhetorical category (Challenge, Background, Approach, Outcome, Future Work). All sentences specifying the purpose for a citation are tagged with the appropriate reason (Criticism, Comparison, Use, Basis, Substantiation or Neutral). Advantages, Disadvantages, Novelties, Common practices or Limitations are identified in any sentence along the document and all of them are graded according to their relevance for a summary.

A collection of handwritten summaries has also been compiled with 3 versions for all the documents in the corpus. This constitutes a valuable resource for training automatic summarization tools.

The data set is downloadable on the web at: http://sempub.taln.upf.edu/dricorpus.

Acknowledgements
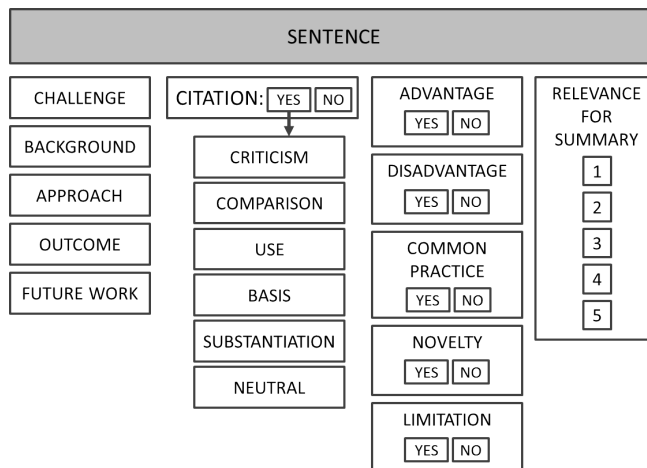
Figure 7: Information contained in every sentence of the Scientific Discourse Annotated Corpus

Bibliographical References

Abu-Jbara, A. and Radev, D. (2012). Reference scope identification in citing sentences. In *Proceedings of the 2012 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, NAACL HLT '12, pages 80–90, Stroudsburg, PA, USA. Association for Computational Linguistics.

Ciancarini, P., Iorio, A. D., Nuzzolese, A. G., Peroni, S., and Vitali, F. (2014). Evaluating citation functions in cito: Cognitive issues. In Valentina Presutti, et al., editors, *ESWC*, volume 8465 of *Lecture Notes in Computer Science*, pages 580–594. Springer.

Fisas, B., Ronzano, F., and Saggion, H. (2015). On the discursive structure of computer graphics research papers. In *The 9th Linguistic Annotation Workshop held in conjuncion with NAACL 2015*, pages 42–51.

Hovy, D., Berg-Kirkpatrick, T., Vaswani, A., and Hovy, E. H. (2013). Learning whom to trust with mace. In *HLT-NAACL*, pages 1120–1130.

Jörg, B. (2008). Towards the nature of citations. In Carola Eschenbach et al., editors, *Poster Proceedings of FOIS 2008*, pages 31–36. DFKI, 10.

Liakata, M., Teufel, S., Siddharthan, A., and Batchelor, C. (2010). Corpora for the conceptualisation and zoning of scientific papers. In Nicoletta Calzolari (Conference Chair), et al., editors, *Proceedings of the Seventh International Conference on Language Resources and Evaluation (LREC'10)*, Valletta, Malta, may. European Language Resources Association (ELRA).

Moravcsik, M. J. and Murugesan, P. (1975). Some results on the function and quality of citations. *Social studies of science*, 5(1):86–92.

Nanba, H. and Okumura, M. (1999). Towards multi-paper summarization reference information. In *Proceedings of the 16th International Joint Conference on Artificial Intelligence - Volume 2*, IJCAI'99, pages 926–931, San Francisco, CA, USA. Morgan Kaufmann Publishers Inc.

Radev, D. R., Teufel, S., Saggion, H., Lam, W., Blitzer,

J., Qi, H., Çelebi, A., Liu, D., and Drabek, E. (2003). Evaluation challenges in large-scale document summarization. In *Proceedings of the 41st Annual Meeting on Association for Computational Linguistics - Volume 1*, ACL '03, pages 375–382, Stroudsburg, PA, USA. Association for Computational Linguistics.

Saggion, H., Radev, D., Teufel, S., Lam, W., and Strassel, S. M. (2002). Developing infrastructure for the evaluation of single and multi-document summarization systems in a cross-lingual environment. In *In LREC 2002, pages 747–754, Las Palmas, Gran Canaria*, pages 747–754.

Spiegel-Rösing, I. (1977). Science studies: Bibliometric and content analysis. *Social Studies of Science*, 7(1):97–113, February.

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

Teufel, S., Siddharthan, A., and Tidhar, D. (2006). Automatic classification of citation function. In *Proceedings of the 2006 Conference on Empirical Methods in Natural Language Processing*, pages 103–110, Sydney, Australia, July. Association for Computational Linguistics.

Teufel, S., Siddharthan, A., and Batchelor, C. (2009). Towards discipline-independent argumentative zoning: Evidence from chemistry and computational linguistics. In *Proceedings of the 2009 Conference on Empirical Methods in Natural Language Processing: Volume 3 - Volume 3*, EMNLP '09, pages 1493–1502, Stroudsburg, PA, USA. Association for Computational Linguistics.

Teufel, S. (2010). *The Structure of Scientific Articles - Applications to Citation Indexing and Summarization.* CSLI Studies in Computational Linguistics. Univ. of Chicago Press.