Meta-discourse markers and problem-structuring in scientific articles

Simone Teufel
Centre for Cognitive Science
University of Edinburgh
S.Teufel@ed.ac.uk

Abstract

Knowledge about the argumentative structure of scientific articles can, amongst other things, be used to improve automatic abstracts. We argue that the argumentative structure of scientific discourse can be automatically detected because reasoning about problems, research tasks and solutions follows predictable patterns.

Certain phrases explicitly mark the rhetorical status (communicative function) of sentences with respect to the global argumentative goal. Examples for such meta-discourse markers are *"in this paper, we have presented ..."* or *"however, their method fails to"*. We report on work in progress about recognizing such meta-comments automatically in research articles from two disciplines: computational linguistics and medicine (cardiology).

1 Motivation

We are interested in a formal description of the document structure of scientific articles from different disciplines. Such a description could be of practical use for many applications in document management; our specific motivation for detecting document structure is quality improvement in automatic abstracting.

Researchers in the field of automatic abstracting largely agree that it is currently not technically feasible to create automatic abstracts based on full text understanding (Sparck Jones 1994). As a result, many researchers have turned to sentence extraction (Kupiec, Pedersen, & Chen 1995; Brandow, Mitze, & Rau 1995; Hovy & Lin 1997). Sentence extraction, which does not involve any deep analysis, has the huge advantage of being robust with respect to individual writing style, discipline and text type (genre). Instead of producing abstracts, this results produces only *extracts*: document surrogates consisting of a number of sentences selected verbatim from the original text.

We consider a concrete document retrieval (DR) scenario in which a researcher wants to select one or more scientific articles from a large scientific database (or even from the Internet) for further inspection. The main task for the searcher is *relevance decision* for each paper: she needs to decide whether or not to spend more time on a paper (read or skim-read it), depending on how useful it presumably is to her current information needs. Traditional sentence extracts can be used as rough-and-ready relevance indicators for this task, but they are not doing a great job at representing the contents of the original document: searchers often get the wrong idea about what the text is about. Much of this has to do with the fact that extracts are typically incoherent texts, consisting of potentially unrelated sentences which have been taken out of their context. Crucially, extracts have no handle at revealing the text's logical and semantic organisation.

More sophisticated, user-tailored abstracts could help the searcher make a fast, informed relevance decision by taking factors like the searcher's expertise and current information need into account. If the searcher is dealing with research she knows well, her information needs might be quite concrete: during the process of writing her own paper she might want to find research which supports her own claims, find out if there are contradictory results to hers in the literature, or compare her results to those of researchers using a similar methodology. A different information need arises when she wants to gain an overview of a new research area – as an only "partially informed user" in this field (Kircz 1991) she will need to find out about specific research goals, the names of the researchers who have contributed the main research ideas in a given time period, along with information of methodology and results in this research field.

There are new functions these abstracts could fulfil. In order to make an informed relevance decision, the searcher needs to judge differences and similarities between papers, e.g. how a given paper relates to similar papers with respect to research goals or methodology, so that she can place the research described in a given paper in the larger picture of the field, a function we call navigation *between* research articles. A similar operation is navigation *within* a paper, which supports searchers in non-linear reading and allows them to find relevant information faster, e.g. numerical results.

We believe that a document surrogate that aims at supporting such functions should characterize research articles in terms of the problems, research tasks and solutions/methodology presented in the specific paper, but it should also represent other researchers' problems, research tasks and solutions mentioned in the paper. Our long-term goal is to automatically reconstruct this problem-solution structure from unrestricted text for the searcher in the form of a *problem-structured* abstract.

But how can we find problems, research questions, research tasks and solutions in text without fully understanding the text? We take sentence extraction as a starting point due to its inherent robustness. Using additional information about each sentence, viz. their rhetorical status with respect to the entire paper, we are in a better position to perform shallow, but guided information extraction, in order to find the information units we are interested in.

In the next section we introduce the level of document structure we are talking about, and the kind of meta-comments we employ in discovering it. In the rest of the paper, we report on ongoing work on automatically filtering meta-comments from annotated and unannotated text. Finding meta-comments in text is an attractive task because it would allow for the automatic adaptation of sytems using such phrases to new domains.

2 Discourse structure and argumentation in scientific articles

Discourse linguistic theory suggests that texts serving a common purpose among a community of users eventually take on a predictable structure of presentation (Kintsch & van Dijk 1978) – and scientific articles certainly serve a well-defined communicative purpose: they "present, retell and refer to the results of specific research" (Salager-Meyer 1992). Particularly in the life and experimental sciences, a rigid building plan for research articles has evolved over the years, where rhetorical divisions tend to be very clearly marked in section headers. Prototypical rhetorical divisions include *Introduction, Purpose, Experimental Design, Results, Discussion, Conclusions.* One of the reasons for this rigidly-defined structure seems to be that the scientific community in these fields has more or less agreed on how to do research: methodologies and evaluation methods are long-lived research entities that do not change often.

One of the corpora we are using is a good example of such texts. It consists of 129 articles in cardiology, taken from the *American Heart Journal*, which have a fixed structure with respect to rhetorical divisions and section headers. The other corpus, in contrast, consisting of 123 (mostly conference) articles in computational linguistics (CL), displays an heterogeneous mixture of methodologies and traditions of presentation one would expect in an interdisciplinary field. Most of the articles cover more than one single discipline, but as a rough estimate one can say that about 45% of the articles in the collection are predominantly technical in style, describing implementations (i.e. engineering solutions); about 25% report on research in theoretical linguistics, with an argumentative tenet; the remaining 30% are empirical (psycholinguistic or psychological experiments or corpus studies). Even though most of the articles have an introduction and conclusions (sometimes occurring under headers with different names), and almost all of them cite previous work, the presentation of the problem and the methodology/solution are idiosyncratic and depend on individual writing style. Very few of the headers in the computational linguistics articles correspond to prototypical rhetorical divisions; the rest contain content specific terminology (cf. Figure 1 which compares relative frequencies of headers for the two corpora).

Computational Linguistics		**Cardiology**	
Header	**Freq.**	**Header**	**Freq.**
Introduction	85%	Introduction	100%
Conclusion	46%	Results	94%
Conclusions	22%	Discussion	94%
Acknowledgments	17%	Methods	92%
Discussion	12%	Tables	79%
Results	11%	Statistics	40%
Experimental Results	9%	Patients	29%
		Limitations	28%
Related Work	7%	Conclusions	25%
Implementation	7%	Statistical Analysis	22%
Evaluation	7%		
Example	6%	Conclusion	17%
Background	6%	Patient Characteristics	9%
Summary	4%		

Figure 1: Highest-frequency headers (with relative occurrence frequencies)

Because the type of research reported in the computational linguistics corpus differs so much, the description of document structure we were looking for had to be flexible enough to generalize over differences in presentation, yet formal enough for the extraction of the information units which are useful for automatic abstracts. We base our model of argumentation in scientific articles on Swales' (1990) CARS model ("Create a Research Space"). Swales' claim is that the main communicative goal of an author of a research article is to convince readers (potential reviewers) that the research described in the paper constitutes an actual contribution to science, in order to have the paper reviewed positively and thus published; this is the case whether or not the paper tries to give the impression that it reports research in an objective, disinterested way. In order to successfully present their case, authors argue in a goal-directed and prototypical way about problem-solving activities — their own and other researchers'. Swales identified prototypical rhetorical building plans of introduction sections, along with linguistics surface cues that signal rhetorical moves. Examples for rhetorical moves include the claim that the paper addresses a *new* problem or, if it is a *well-known* problem, then the presented solution has to be better than that of other researchers.

A first analysis of the corpora confirmed many of the rhetorical building blocks suggested by Swales. We adapted Swales' scheme to the one shown in Figure 8 (at the end of this paper). In the medical corpus, almost all of the moves we found were of type I (*Explicit mention*); the rigid document structure seems to have replaced much of the "argument about problem-solving activities" (types II-V) for which we found ample evidence in the computational linguistics corpus.

We are interested in identifying these moves automatically and shallowly in text, and we believe that this is technically feasible, because the stereotypical, predictable overall structure of the argument can be exploited in doing so.

3 Meta-discourse markers

In this paper we focus on the linguistic realisations of rhetorical moves, i.e. the surfacy signals of the argumentative status of a given sentence. Consider the strings in boldface on the right hand side of Figure 8. They cover the activities of reporting about research (reporting and presenting verbs), the problem-solving process (problems, solutions, tasks); they also include other semantic links like necessity, causality and contrast. Due to explicit or implicit argumentation, many of these strings are evaluative (*"efficient, elegant, innovative, insightful"* vs. *"impossible, inadequate, inconclusive, insufficient"*). We call them *meta-comments* because they talk *about* information units, as opposed to being subject matter (scientific content). Such meta-comments are very frequent in our collection.

Our meta-comments are similar Paice's (1981) *indicator phrases* (he was the first to use such phrases for abstracting); they are less similar to *cue phrases*, the discourse markers usually studied in discourse analysis, because they are *not* sentence connectives (with some exceptions), and because they are typically considerably longer and far more varied.

The fact that the computational linguistics texts stem from an unmoderated medium (i.e. they are neither chosen for publication nor edited by a central authority), means that there were no external restrictions on how exactly to say things. Authors use idiosyncratic style, which can vary from formal to informal. There are meta-comments that tend to get used in a fixed, formulaic way, but interestingly, we observed a wide range of linguistic variability with respect to the realization of some of the meta-comments (whereas their semantics is usually perfectly unambiguous). This effect makes the meta-comments in this text type interesting linguistic objects to study.

We observed that there are certain meta-comments which are restricted to certain moves, mostly the evaluative and contrastive phrases and the phrases occurring in moves of type I (*Explicit mention*). Others occur frequently across moves, particularly general argumentative phrases and relevance markers such as *"important"*, *"in this paper, we"*. Argumentative phrases like *"we argue that"* appeared with solution and problem-related moves almost as often as with claims and conclusions. These phrases seemed to be the ones that were most formulaic/fixed across texts.

Our goal is to automatically find meta-markers and associate them with rhetorical moves (where this makes sense). In the next section, we report on a first experiment in that direction.

4 Our experiment

If it is true that most meta-comments are formulaic, recurring expressions, then frequency information should help us separate meta-comments from domain-specific parts of the sentence. Those strings which occur rarely in the corpus will most likely be domain-specific and will appear low on frequency listings of strings, whereas meta-comments should appear high on the lists.

We also used a lexicon of 433 *lexical seeds*. Lexical seeds are words which are semantically related to the activities of reporting, problem-solving, arguing or evaluating, or expressions of deixis (*"we..."*) or other textual cues (e.g. literature references in text were marked up using the symbol [REF], which is a signal for mentions of other researchers' solutions, tasks or problems).

The computational linguistics corpus was drawn from the computation and language archive (`http://xxx.lanl.gov/cmp-lg`) and contains 123 articles; the 129 articles of the cardiology corpus appeared in the *American Heart Journal*. The medical corpus is smaller in overall size (436,909 words vs. 654,477; 14,770 sentences vs. 23,072).

For the computational linguistics corpus, we additionally had a collection of 948 sentences that had been identified as relevant by a human annotator in a prior experiment (Teufel & Moens 1997). A human judge annotated these with respect to the 23 rhetorical moves introduced in Figure 8.

4.1 Filtering

First, we compiled the two corpora into those bigrams, trigrams, 4-grams, 5-grams and 6-grams which did not cross sentence boundaries. We worked with a short stop-list compiled from the corpus (60 highest-frequent words) from which we had excluded those which we expected to be important in an argumentative domain, e.g. personal and demonstrative pronouns. We lowercased all words and counted punctuation (including brackets) as a full word.

We then filtered the n-grams through our seed lexicon, i.e. we retained those expressions which contain at least one of the words of the seed-lexicon (or a morphological variant of it). We also compiled and counted n-grams for the 948 computational linguistics target sentences, to see how similar these phrases from the annotated parts were to the filtered or unfiltered bigrams from the entire corpus.

Before filtering		After filtering	
354	[ref] , [ref] ,	118	in this paper ,
301	[ref]) .	111	in ([cref])
297	, [ref] , [ref]	110	can be used to
178	[cref]) .	106	in figure [cref] .
144	[ref] , [ref] .	100	([cref]) ,
139	on the other hand	99	on the basis of
134	for example , the	83	shown in figure [cref]
118	in this paper ,	83	in section [cref] .
116	the other hand ,	75	this paper , we
111	in ([cref])	75	in section [cref] ,
110	can be used to	71	it is possible to

Figure 2: 4-grams in entire CL corpus

The list of 4-grams for the computational linguistics corpus shows a typical picture of the outcome of this process (Figure 2). Before filtering, the frequent corpus n-grams contain general comments and expressions like *"for example"* but content specific expressions are already filtered out. (Very rarely, there are some content-specific phrases like "*natural language*" in the lists — this is due to the fact that the corpus, even though interdisciplinary in nature, is composed of papers focusing on language.) After filtering, the meta-comments on the lists have two properties: (a) they are frequent (b) they contain lexical items that *could* be related to argumentation about problems, research tasks, solutions — in the computational linguistics corpus, these conditions seem to be enough to produce expressions that are good candidates for meta-comments. Unfortunately, condition (a) means that a large number of meta-comments were lost, because they were of low frequency.

Target sentences			
49	in this paper ,	10	can be used to
37	this paper , we	9	in this paper is
25	[ref] , [ref] ,	7	[ref] , [ref] .
22	, [ref] , [ref]	7	described in this paper
18	in this paper we	7	this paper , i

Figure 3: 4-grams in CL target sentences

How similar are the lists for annotated text and entire corpus in the computational linguistics domain? Table 3 shows that they look very similar apart from minor differences, e.g. the fact that the list gained from annotated data contains more [CREF] items (internal cross reference like *"section [CREF]"*) which tend to appear frequently in the sentences where authors state the organisation of the paper. The expressions used in such organisation statements are typically formulaic and recurrent, but not many of these sentences were considered relevant when the 948 target sentences were determined.

Before filtering		After filtering	
120	p < 0.05)	85	on the basis of
102	p < 0.01)	66	of this study was
93	left ventricular ejection fraction	64	in this study ,
86	p < 0.001)	58	this study was to
86	p < 0.0001)	57	patients with heart failure
85	on the basis of	45	the purpose of this
72	at the time of	44	there were no significant
66	of this study was	40	new york heart association
64	in this study ,	39	purpose of this study
59	95 % confidence interval	35	there was no significant
58	this study was to	32	were no significant differences
55	coronary artery disease .	31	p = not significant
49	acute myocardial infarction .	30	has been shown to

Figure 4: 4-grams in medical corpus

The medical corpus shows significant differences (cf. Figure 4). Firstly, unfiltered bigrams do not separate content matter from meta-discourse. In these lists, there are phrases pertaining to statistical analyses (*"$p < 0.01$"*) and several domain-specific phrases. Filtering (right hand side of Figure 4) forces the few meta-comments that *are* being used to the top of the list; they are linguistically invariant. For instance, "*study*" seems the only acceptable expression used for the current research, whereas the range is much wider in the other corpus (*"paper, article, study, work, research..."*, and all the meta-comment candidates in the top part of the list belonged to one single rhetorical move, viz. *Explicit Mention of the Research Task* (Ex-T in Figure 8).

A certain amount of noise has been introduced through the seed-lexicon because word senses were not disambiguated: *"failure"* was included in the seed lexicon to indicate mentions of failure of other researchers' solution. Because this term obviously has the different meaning of *"heart failure"* in the cardiology context, the desired distinction between subject matter strings and meta-comments got lost; similarly *"New York"* was included because the word *"new"* could potentially point to novel approaches. This might mean that it is necessary to use different stop-lists and/or seed lexicons for different domains.

As we have seen before, associating meta-comments with rhetorical moves is a more difficult task for some meta-comments than for others. We tried to anchor the probable rhetorical move of a phrase in the lexical seed it contains, a simplification we are forced to make due to the small amount of annotated text we have available (which is reflected in the low numbers). We are thus in the process of working on a larger scale annotation.

We used the human judgements to count how often each word contained in the target sentences appears with a certain rhetorical move. If the difference in frequency between the best-scoring moves for that word was large enough, we assumed it was a good indicator for the highest-scoring move, and we then manually associated the given rhetorical move with the word if it was contained in the seed lexicon, or to semantically similar seeds. For example, seeds that are the most likely associated with the OWN SOLUTION BETTER (53 examples of this move in the target sentences) were *"than"* (39), *"better"* (36), *"results"* (21), *"method"* (19), *"using"* (15), *"significantly"* (14), *" outperforms"* (12) and *"more"* (12). Filtered meta-comments are then assigned the rhetorical move predicted by the first seed they contain. Figure 5 shows the meta-comments filtered for the seed *"better"* from both corpora. In the medical corpus, there is less argument about methodology/solutions, and as a result the phrases found are unfortunately *not* meta-comments but contain medical terminology.

Computational Linguistics		Cardiology	
64	better than	11	a better
50	a better	7	better than
23	better than the	6	to better
20	much better	6	significantly better
19	the better	6	better in
19	is better	5	better left
16	significantly better	5	better left ventricular
16	be better	5	and better
13	better performance	4	better preserved
11	are better	4	better in smokers
10	better the	3	to be somewhat better tolerated
9	significantly better than	3	failure symptoms in spite of better
6	better suited	3	better preserved left ventricular systolic function
6	better than that of	3	better in smokers than in nonsmokers

Figure 5: Potential meta-comments with *"better"* from both corpora

Also, we observed that it is not easy to predict the optimal length of a certain meta-comment which is indicative of a certain rhetorical move. For moves containing *other problems/solutions/tasks* the very short string *"[REF]"* is contained in all successful meta-comments, whereas for explicit mention of research goals, the maximal length 6 of meta-comments which we chose for these experiments might even be too short. As another example, for the STEP move (*"goal is achieved by doing solution"*), the best indicator we found was *"in order to"*.

4.2 Evaluation

We evaluate the quality of these automatically generated meta-comment lists by comparing them to a manually created meta-comment list used by a summarisation system, cf. (Teufel & Moens To Appear). The performance of the system – with the two different meta-comment lists – is measured by precision and recall values of co-selection with the target extracts defined by human annotators mentioned earlier. The summarisation process consists of two consecutive steps, sentence extraction and rhetorical classification, and uses other heuristics like location and term frequency.

The summarisation system requires a list of meta-comments of arbitrary length, containing a *quality score* for each phrase which estimates how predictive these phrases are in pointing to extract-worthy sentences, and the most likely rhetorical label that sentences with this meta-comment will receive.

Quality Class	Rhetorical Move	Meta-comment
2	–	paper ,
3	–	this paper presents a
2	STEP	in order to
3	–	in this paper , we will
2	EX-T	in this paper we have
1	CO-S	unlike [ref]
1	EX-T	this paper is to
3	EX-T	in this paper , we describe
2	EX-P	paper is
2	–	paper we
1	EX-T	this paper has presented
1	EX-T	, we propose a method
1	–	in passage ([cref]
1	–	, we argue that
1	–	argue that
2	EX-T	method for
1	EX-C	we show that the
1	EX-T	show how
1	–	property and the number
1	EX-T	the advantages of
1	EX-E	the wall street journal
1	NEC-S-T	the importance of
1	CO-C	however , we
1	EX-T	be used to

Figure 6: Extracts from automatic list of meta-comments

We automatically built the meta-comment list in Figure 6 (containing 318 entries). We started from all n-grams compiled from the target sentences and took the following heuristics into account: Firstly, choose phrases with a high ratio of target frequency to corpus frequency, because these are *indicative* phrases. Set the quality value accordingly. Secondly, exclude phrases with a low overall frequency, or decrease their quality score, because including/overestimating them might construct a model that is over-fitted to the data. Thirdly, associate each phrase with its most likely rhetorical move, by taking the ratio between frequency in each rhetorical class and the frequency of the rhetorical label itself into account. If below a certain threshold, don't associate any move at all (e.g. *"paper ,"* in Figure 6).

The manual meta-comment list, in contrast, was compiled in an extremely labour intensive manner and refined over the months. It consists of 1791 meta-comments (some of which are much longer than the maximum of 6 words that the automatic phrases consisted of), along with their most plausible rhetorical moves and quality scores.

	Manual	**Automatic**
Extraction	66.4%	52.5%
Classification	66.3%	54.3%

Figure 7: Evaluation results: precison and recall of co-selection

As Figure 7 shows, using the automatic meta-comment list instead of the manually created one decreased the summarizer's performance from 66.4% to 52.5% precision and recall for extraction, and from 66.3% to 54.3% precision and recall for classification.

5 Discussion and further work

The evaluation indicates that the quality of the automatic meta-comment list is not yet high enough to replace the manual list in our summarization system. However, a look at the automatic list itself shows that, even though it is far from perfect, most of the high-frequent strings found are plausible candidates for meta-comments (or parts of meta-comments). In most cases, subject matter can be successfully filtered out.

We regard the simple method for automatic meta-comment identification discussed in this paper as a baseline for further work. We have simplified the problem of finding meta-comments enormously by only considering verbatim substrings. By doing so, we have ignored discontinuous strings, morphological variation and statistical interaction between the words in the string. In addition, the phrases considered so far have been short, and we have not collected many of them, because we wanted to rely only on the ones with reasonably high frequencies.

The biggest problem for now is that highly indicative, but infrequent meta-comments cannot be found with a simple method like ours. Therefore, it is essential to perform some generalization over similar phrases. One way would be the automatic clustering of similar concepts, e.g. to find out that *"argue"* and *"show"* are presentational verbs with similar semantics. Another idea would be to allow for more flexible patterns consisting of short n-grams and other words, in order to skip over intervening words like adjectives and adverbs. This might avoid the data sparseness problems encountered with the longer n-grams.

6 Summary

We have presented some baseline results from our ongoing work concerning automatic filtering of meta-comments (indicator phrases) from scientific papers. Meta-comments can vary considerably from one domain to another, as the comparison of the two corpora we considered shows. In the computational linguistics articles, authors argue explicitly about problems, solutions and research tasks, whereas this is less the case in the medical domain, where meta-comments are less frequent and more formulaic.

We have shown that lists of meta-comments acquired in a simple automatic process can be used to automatically identify a shallow document structure in scientific text, albeit with a certain quality loss when compared to manually constructed resources.

7 Acknowledgements

Data collection of the computational linguistics corpus took place collaboratively with Byron Georgantopolous. We thank Kathy McKeown, Vasileios Hatzivassiloglou and Olga Merport from Columbia University, NY, for kindly allowing us to use their cardiography corpus.

References

Brandow, R.; Mitze, K.; and Rau, L. F. 1995. Automatic condensation of electronic publications by sentence selection. *Information Processing and Management* 31(5):675–685.

Hovy, E. H., and Lin, C. Y. 1997. Automated text summarization in summarist. In Mani, I., and Maybury, M. T., eds., *Proceedings of the workshop on Intelligent Scalable Text Summarization, in association with ACL/EACL-97.*

Kintsch, W., and van Dijk, T. A. 1978. Toward a model of text comprehension and production. *Psychological Review* 85(5):363–394.

Kircz, J. G. 1991. The rhetorical structure of scientific articles: the case for argumentational analysis in information retrieval. *Journal of Documentation* 47(4):354–372.

Kupiec, J.; Pedersen, J. O.; and Chen, F. 1995. A trainable document summarizer. In *Proceedings of the 18th ACM-SIGIR Conference, Association for Computing Machinery, Special Interest Group Information Retrieval*, 68–73.

Salager-Meyer, F. 1992. A text-type and move analysis study of verb tense and modality distributions in medical English abstracts. *English for Specific Purposes* 11:93–113.

Sparck Jones, K. 1994. Discourse modelling for automatic summarising. Technical report, Computer Laboratory, University of Cambridge. TR-290.

Swales, J. 1990. *Genre analysis: English in academic and research settings.* Cambridge University Press.

Teufel, S., and Moens, M. 1997. Sentence extraction as a classification task. In *ACL/EACL-97 Workshop 'Intelligent Scalable Text Summarization'*, 58–65.

Teufel, S., and Moens, M. To Appear. Argumentative classification of extracted sentences as a first step towards flexible abstracting. In Mani, I., and Maybury, M., eds., *Advances in automatic text summarization*. MIT Press.

I. Introduction of problems, tasks, solutions by explicit mention		
Ex-T	Own research task	**The aim of this paper is to** *examine the role that training plays in the tagging process.*
Ex-S	Own solution	**The basic idea for the analysis** *can be seen as a logical counterpart at the glue level of the standard type assignment for generalized quantifiers [REF].*
Ex-O-S	Other solution	**The traditional approach has been to** *plot isoglosses, delineating regions where the same word is used for the same concept.*
Ex-C	Own conclusions/ claims	*In our corpus study,* **we found that** *three types of utterances (prompts, repetitions and summaries) were consistently used to signal control shifts.*
Ex-O-C	Other conclusions/ claims	**It has often been stated that** *discourse is an inherently collaborative process.*
Ex-E	Own evaluation methodology	*In this section* **we evaluate the performance of** *the methodology implemented* **by** *predict***ing** *succeeding words by using preceding words.*
Ex-R	Own (numerical) results	**The evaluation of the accuracy of** *the rhetorical structure analysis carried out previously ([REF])* **showed** *74* **%.**
II. Contrastive introduction of problems, tasks, solutions		
Co-S	Contrast between own and other solutions	*In this paper, we argue that* **instead of** *apply***ing** *the arbitration process to the discourse level,* **it should** *be applied to the beliefs proposed by the discourse actions.*
Co-T	Contrast between own and other tasks/ problems	**Unlike most research in** *pragmatics that focuses on certain types of presuppositions or implicatures,* **we provide** *a global framework in which one can express all these types of pragmatic inferences.*
Co-C	Contrast between own and other claims	**Despite the hypothesis that** *the free word order of German leads to poor performance of low order HMM taggers when compared with a language like English,* **we have shown that** *the overall results for German are very much along the lines of comparable implementations for English, if not better.*
III. Attribution of properties to problems, tasks, solutions		
Imp-T	Own research task is important	**The last decade has seen a growing interest in** *the application of machine learning to different kinds of linguistic domains.*
Hard-T	Own research task is hard	**One of the difficult problems in** *machine translation from Japanese to English or other European languages* **is** *the treatment of articles and numbers.*
New-T	Own problem/ research task is new	**No** *formal framework* **has been proposed, to our knowledge, to** *regulate the interaction between regular and exceptional grammatical resources.*
Good-S	Own solution is advantageous	*First, it is in certain respects* **simpler**, *in that* **it requires no** *postulation of otherwise unmotivated ambiguities in the source clause.*
Bad-O-S	Other solution is flawed	*However, we argue that such formalisms* **offer little help to** *computational linguists in practice.*
IV. Functional relations between problems, tasks, solutions		
Solves	Own solution solves own research task	**This account** *also* **explains** *similar differences in felicity for other coordinating conjunctions as discussed in [REF].*
Avoids	Own solution avoids problems	*We have introduced a simple, natural definition of synchronous tree-adjoining derivation* **that avoids the** *expressivity and implementability* **problems** *of the original rewriting definition.*
Step	Own solution is a step towards research task	**We have** *thus* **developed** *an evaluation heuristic that combines several different measures,* **in order to** *select the parse that is deemed overall "best".*
Nec-S-T	Own solution necessary to achieve research task	*We have argued that obligations* **play an important role in** *accounting for the interactions in dialog.*
NotSo	Other solution does not solve problem/ task	*Dependency grammar* **runs into substantial difficulty trying to account for** *the proform one.*
New-P	Other solution introduces new problems	*Specifically,* **if a treatment such as [REF]'s is used** *to explain the forward progression of time in example [CREF],* **then it must be explained why** *sentence [CREF] is as felicitous as sentence [CREF].*
V. Direct comparison of problems, tasks, solutions		
Better-S	Own solution is better than other solution	*We found that the MDL-based method* **performs better than** *the MLE-based method.*
Harder-T	Own research task is harder than other task	*. . . disambiguating word senses to the level of fine-grainedness found in WordNet* **is quite a bit more difficult than** *disambiguation to the level of homographs [REF], [REF].*

Figure 8: Rhetorical moves in scientific papers; examples from our corpus of computational linguistics