Content selection in summary generation

Lucia Helena Machado Rino[1]
and
Donia Scott
Information Technology Research Institute
University of Brighton
Lewes Road, Brighton BN2 4AT
UK
Lucia.Rino,Donia.Scott@itri.bton.ac.uk

Abstract

We report on work aimed at providing discourse strategies for the automatic generation of draft scientific summaries. We have developed a discourse model based upon the theory of discourse structure and the analysis of data corresponding to naturally produced summaries written by domain-expert writers.

In this paper, we focus on the possible ways of compression of discourse structures that convey a certain degree of detail. We present some heuristics for compression of summaries, which take into account those discourse relations that link at least one optional discourse segment. When applied to the data, the heuristics provided coherent and clear summaries, significantly shorter than the originals. These heuristics guide the process of compression in the automatic summary generator that is currently under development.

Introduction

The work described in this paper is part of a project for the automatic generation of summaries of scientific papers in English. Summarisation involves, among other things, the selective choice of key information units to convey to the user.

In this paper, we will report on some of the techniques for driving automatic content selection. Discourse strategies select and organise information for the generation of different versions of summaries. Only content units that are significant for a particular readership must be expressed in the generated summaries, where by significant we mean being understandable and relevant for the reader's purposes. The versions differ on parameters of informativeness and conciseness, for different communicative purposes.

The source of information for the generation of summaries is a set of content units (e.g. propositions) that, when selected and organised coherently in a discourse structure, can be expressed in textual form by means of cohesive devices. The units are classified according

[1]On leave from the Universidade Federal de São Carlos - SP - Brazil. This work is supported by the National Council for Scientific and Technological Development (CNPq), Project No. 201610/92-2, and the Fapesp Project No. 92/2151-8.

Proceedings of the XI Simposio Brasileiro de Inteligecia Artificial, Fortaleza, Brazil, October 1994

to their relevance, and they can be optional or obligatory (i.e. non-essential or essential for the message to be conveyed).

Different strategies can be employed for the selection and organisation of different content units. For example, for two types of summaries, informative (those which present substantive information of the corresponding paper) and indicative (those which have a referential function, signalling the content of the corresponding text), information is conveyed differently (Hutchins, 1987; Paice, 1990). Informative summaries must convey most information available; indicative ones must omit them. The message transmitted using these strategies must be always clear and coherent.

We assume that a discourse structure of an informative summary carries optional content units that are candidates for suppression, and thus the compression of the structure is more likely to lead to an indicative summary. The compression is possible when discourse relations (which apply recursively between pairs of optional or obligatory units of information) link at least one non-essential content unit.

In this framework, the interplay between clarity and conciseness is very complex: a clear summary for a domain-expert reader may be obscure for a naïve reader. A more concise summary may be motivating even if it does not convey the same amount of information a less concise version does. Clarity is closely related to factors such as informativeness, explicitness, signalling of discourse, and amount of detail.

We present in this paper an investigation of a specific aspect of the interdependence between clarity and conciseness: the degree of informativeness in the selection and organisation of information for the generation of indicative or less detailed summaries, based upon informative discourse structures. A discourse model has been developed through the study of a corpus collected via an empirical study consisting of a summarisation task in the domain of physics, carried out by domain-expert writers. We will present some of the heuristics derived for guiding content selection for the different types of summary. We first report on the adopted approach, and then present the heuristics that provide the necessary compression. Throughout we will illustrate the operation of the heuristics with respect to the following naturally produced summary, shown here with clause delimitation:

Text 1: Original summary delimited by clauses[2]

1. The interaction of continuous CO2 laser radiation focused onto a free water surface is studied, both in normal gravity and in reduced gravity conditions.
2a. The [2b] depth of the keyhole structures produced by different laser powers are found to be in good agreement with the theory of Andrews and Atthey.
2b. observed
3a. This theory includes the recoil pressure of the evaporation, hydrostatic pressure and surface tension
3b. but does not include dynamic effects,
3c. such as the momentum reaction flow.
4. The shape of the keyhole and variation of depth with gravity are also calculated.
5a. The size distribution of bubbles produced at the tip of the keyhole has been measured

[2]Clauses represented in square brackets appear in the natural flow of the sentence, and they are generally expressed as adjectives, or adjectival clauses.

5b. and the [5c] is explained by the increasing sharpness of the tip.
5c. observed trend towards smaller bubbles at higher power
6a. Using analysis of the balanced forces on bubbles trapped under the keyhole,
6b. the speed of the momentum reaction flow down the side of the keyhole has been calculated to be about 20 cm/s.
7. This is a significant flow which has not been considered in previous theoretical models.
8a. Very large bubbles have been observed during a transition into low-gravity conditions
8b. and are partially explained in terms of the pressure difference between the narrow keyhole and the initial bubble.
9. It is suggested that this effect may be of significance in laser-beam welding.

(217 words)

Linguistic analysis

Recent research in computational linguistics, and particularly in summarisation, have focused on theories of discourse structure for the representation of the discourse under investigation, in the search for global discourse structures that can be the legitimate way for automation. Most of the current approaches rely on some kind of global representation of discourse for particular genres and styles of text (Endres-Niggemeyer, 1993; Liddy, 1991; McKeown, 1993; Paice, 1990; Sparck-Jones, 1993b). Some suggest schematic representations of discourse strategies (Liddy et al., 1993; McKeown, 1985; Paris, 1993) for the generation of texts.

Much of the work developed so far for summarisation addresses discourse modelling by means of analysing corpora linguistically. Often, the surface signalling of discourse structure (i.e. lexical and syntactical clues) is the focus of the analysis, along with studies of the order of components, likelihood of component occurrence, distribution of verb tense, and continuation clues (Francis and Liddy, 1991; Liddy, 1991; Paice, 1990; Sparck-Jones, 1993a). This approach is based on discourse theory (Hoey, 1983; Hutchins, 1977; Hutchins, 1987; Kintsch and van Dijk, 1978), and constitutes a meaningful contribution for the specification of essential information for automatic generation.

In this work, we take a similar approach, by analysing linguistically the corpus of physics summaries in order to model discourse for summarisation. Schematic representations of discourse strategies will be derived for content selection and organisation of information. The analysis aims at reconstructing the summaries' discourse structures. In this process, a pattern of scientific discourse known as Problem-Solution (P-S) was adopted (Hoey, 1979; Hoey, 1983; Hutchins, 1977).

At a macro-structural level, the summaries can be expressed by the sequence

Situation-Problem-Solution/Response-Results-Evaluation

These macro-components can be further detailed in terms of other macro-components. For example, Results can be linked to their particular evaluation.

The sequence of macro-components corresponds to a semantic progression of discourse segments that are connected by a logical deductive schema (Hutchins, 1977). This schema has proved to be useful in the linguistic analysis, because it allows the recognition of the coherent thread of discourse, by organising thematic and semantic progression at all levels of text structure[3].

A much more intricate logical sequence can take place in scientific discourse. Figure 1 can roughly represent the possible sequences ("Begin" and "End" tags highlight initial and final structure organisers, respectively[4]).

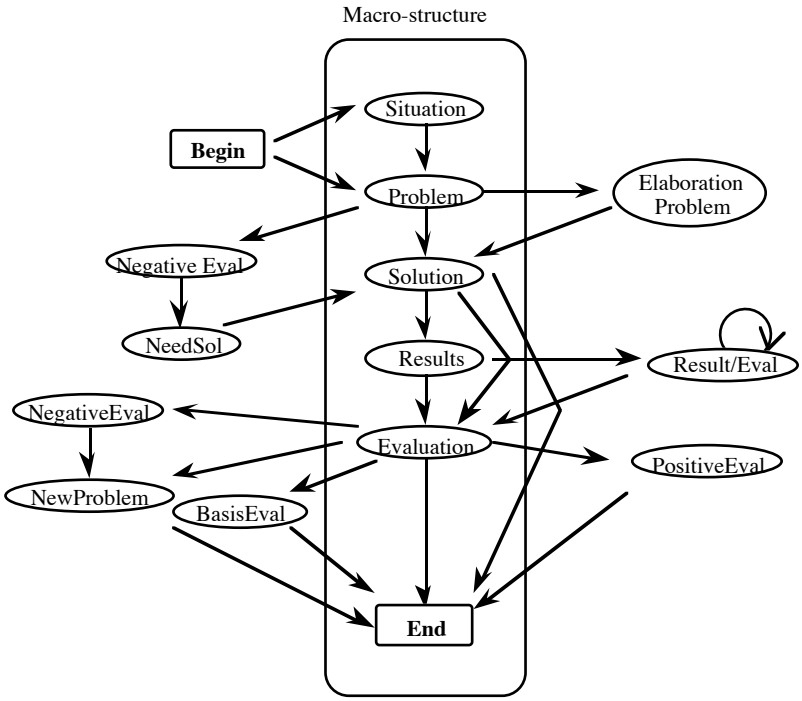

Figure 1: The Problem-Solution structure

In the P-S paradigm, the following assumptions hold:

1. Discourse is organised.
2. This organisation is, at least in part, hierarchical.
3. The writer can assess the well-formedness of a discourse.
4. Aspects of the surface of the discourse make readers and writers perceive its underlying structure. For readers, the surface contains enough clues for the perception of an

[3]Although such a schema was first associated to full texts, it is also recognisable in the analysed condensed forms.

[4]The minimum path comprises only the discourse components Problem and Solution.

accurate discourse organisation (this is the basis for comprehension); for writers, the discourse structure must be mapped onto language on the basis of the competence of the readers, so that they can reconstruct the writer's intended meaning.

5. It is possible to build infinite patterns of discourse out of a finite number of resources.

6. At lower levels of discourse (commonly associated to paragraphs) there are a finite number of possible relations.

7. Discourse is generally represented as a logical sequence of propositions. In its natural form, the sequence is unmarked (it is logically implicit). When altered, linguistic markers must be introduced (e.g. subordination and conjuncts), so that the sequence is directly and uniquely recognisable.

Discourse relations

In the linguistic analysis carried out, discourse relations have been assumed to exist between clauses or groups of clauses. Mann and Thompson's (1987) claim that "the same sorts of relations characterize text structure at all levels." (page 40) has been considered, in order to recognise the links between different levels of discourse organisation.

Relations such as Cause-Consequence (Ca-Co), Condition-Consequence (Cond-Cons), Instrument-Achievement (I-A), and Instrument-Purpose (I-P), according to Hoey, can be mapped onto clause relations. For instance, a Ca-Co holds between the statement of a Problem and the statement of a Solution: "Because there is a problem, a solution is searched for."

Building on the work of Hoey, we also make use of other discourse relations to express the links between discourse segments at more fine-grained levels. For example, we include in our framework the relations Purpose, Enablement and Justification (Mann and Thompson, 1987); Means (Moore, 1989); Exemplification and Explanation (Hobbs, 1985); Background, Evaluation and Elaboration (Hobbs, 1985; Mann and Thompson, 1987). According to our corpus, some of these relations hold only at the macro-level, others allow a more fine-grained representation. In addition, some relations address both levels of discourse representation. Those relations that are too abstract for a good characterisation of the discourse organisation are further detailed. For example, General-Particular relations can be expressed by General-Example and Preview-Detail relations. General-Example relations can be clearly represented as Hobbs' *Exemplification* coherence relation (Hobbs, 1985). The necessary modificants will not be reported here.

Texture[5] is considered to hold by assuming that clauses are related to each other by means of two broad classes of relations:

Logical Sequence They hold between successive events or ideas and have functional meaning in themselves, because they express the logical deductive reasoning.

Matching relations They assure connectedness between text components and very often permeate the Logical Sequence relations.

[5]Texture concerns the organisation of material, as an overlaying of patterns created by the readers' perceptions of the content, the rhetorical structure of the message, and the linguistic structure of the physical text (Stoddard, 1991).

We have determined that matching relations have at least two different functions. The first is to guarantee coherent and unambiguous representations at a deeper level of discourse organisation (e.g. determining thematic functions in order to articulate discourse relations). The second is to map structure onto language, in which case they are associated to grammatical features (e.g. repetition of a noun group by lexical substitution).

General-Particular relations, along with Matching ones, are the organisers of discourse for comprehension, at a more fine-grained level. Very often, the former can also be expressed by means of the latter.

The importance of these two categories of discourse relations in the linguistic analysis is twofold:

- They ease the process of recognition of relations between discourse segments, because they are mapped onto the surface of the discourse by means of grammatical constructions. They relate text spans through explicit signals of the discourse structure, or through implicit semantic constructions (e.g. subordinations, conjunctions).

- They aggregate information in a coherent progression of matching devices. For example, they introduce an example, or more detail about a previous mentioned entity of discourse.

The reconstruction of the discourse structures of the analysed summaries was primarily based on such relations. The relations also guide the definition of the heuristics for compression presented in this paper.

Definition of heuristics for compression of a discourse structure

We adopt the following assumptions for compression in summary generation:

- In the discourse structure under investigation, there are non-essential units of information that can be omitted.

- Whenever omitting a discourse component from a discourse structure, the resulting text is still understandable, at least for more knowledgeable readers (i.e. those able to bridge the resulting inference gaps).

- In omitting discourse components, all the possible referents are resolved in advance.

- The deletion of complex discourse segments (i.e. those that do not relate only propositional-like units) implies the deletion of the related sub-structures.

- The suppression of intermediary relations of a discourse structure does not necessarily imply the suppression of its macro-components.

- Omitting all the allowable (i.e. optional, or feasible for omission according to the heuristics) discourse components of a discourse structure leads to a highly indicative summary.

- The same discourse component that can be omitted in one context may be obligatory in another, depending on the addressed readership[6].

- The quality of the summaries resulting from the application of some heuristics is not evaluated at this stage of the research. However, coherence must be assured, even though it may not be explicitly marked by cohesive devices.

The structural and content variants extracted from the corpus show an interaction between selection and organisation of information for compression: some discourse relations are recognisable as holding between optional units of information, and can be considered for the generation of indicative summaries by omitting some or all of the components they relate (in this case, the relations themselves can be deleted).

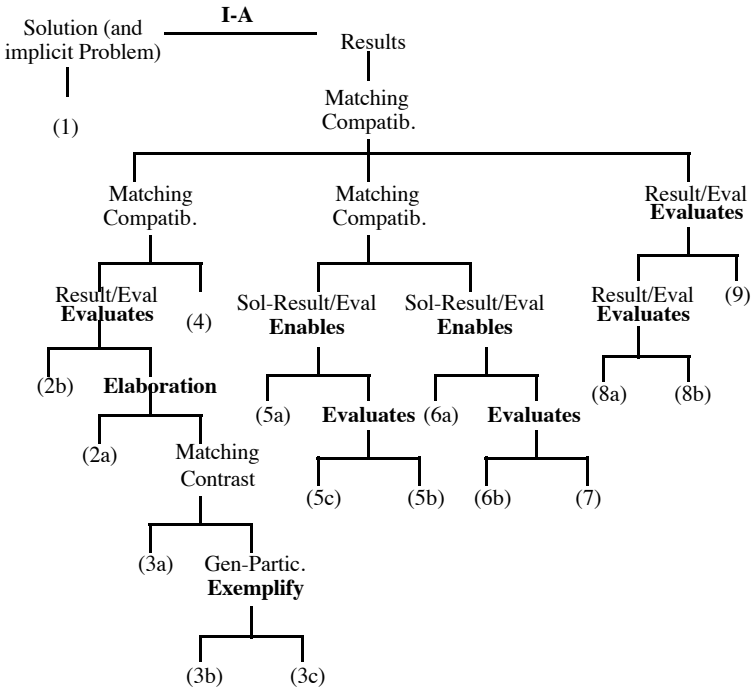

Figure 2: Discourse structure of a naturally-produced summary

Compression can happen in two ways: (a) by considering the suppression of macro-components of the discourse structure; (b) by considering the suppression of information related to the expression of more detail (which features specific discourse relations).

In case (a), eliminating a macro-component of the discourse structure implies in eliminating all its derived sub-structures (this is one of the assumptions).

[6]We notice that, although we are not addressing in this paper the selection of content according to the readership, such an assumption allows different content selection to take place for different versions of summaries.

In case (b), eliminating details implies in analysing which relations that hold in the intermediate level of discourse organisation can be suppressed, with no prejudice for coherence. More detail is directly linked to General-Particular and Matching relations in the P-S paradigm. Other relations, such as Elaboration, which can be further refined as Part-Whole, General-Specific, Abstract-Instance, and Attribute (Hovy, 1990), also allow compression (they can be used for generalisation, which is a widely used device for summarisation).

In what follows, we outline some of the heuristics we have developed for content selection and give examples of their application with reference to the naturally produced summary shown in Text 1. The discourse structure of the exemplified summary is shown in Figure 2.

> **Hypothesis 1:** In General-Particular relations, the *General* discourse segment is a generalisation of the *Particular* discourse segment.
>
> **Heuristic 1**: Delete Particular, from General-Particular relations.

Applying this heuristic to clauses 3b-3c of Text 1, given by:

> `[This theory]` (*resolution of the reference to clause 3a*) `does not include dynamic effects, such as the momentum reaction flow.`

the resulting sentence is

> `This theory does not include dynamic effects.`

> **Hypothesis 2:** The statement of Problem and Solution is enough for a domain-knowledgeable reader to understand what the reported research is about.
>
> **Heuristic 2**: Delete every discourse segment that is optional, and linked at any level to Problem-Solution segment[7].

When this heuristic is applied to the entire summary, the result is the minimum summary

> `The interaction of continuous CO2 laser radiation focused onto a free water surface is studied, both in normal gravity and in reduced gravity conditions.`
>
> (24 words)

> **Hypothesis 3:**An elaboration corresponds to information added to a specific component of a discourse segment, for clarification.
>
> **Heuristic 3**: Delete Y, in *Elaboration(X,Y)*

[7]A similar method has been suggested for summarisation purposes: if we consider only the most significant units of information linked to the macro-components (e.g. the topic information, or the most nuclear, in rhetorical terms), the result will be a minimum skeleton, corresponding to a highly indicative summary (Hoey, 1983).

Applying this heuristic to the sentences 2-3 of Text 1, given by:

> `The observed depth of the keyhole structures produced by different laser powers are found to be in good agreement with the theory of Andrews and Atthey. This theory includes the recoil pressure of the evaporation hydrostatic pressure and surface tension but does not include dynamic effects, such as the momentum reaction flow.`

the resulting sentence is

> `The observed depth of the keyhole structures produced by different laser powers are found to be in good agreement with the theory of Andrews and Atthey.`

> **Hypothesis 4:** Indicative summaries do not present specific results, since these correspond to substantive information.
>
> **Heuristic 4**: Delete Results, in *I-A(Solution,Results)*

Applying this heuristic to the illustrated summary leads to the very same solution as heuristic 2. The reason for this is that the discourse structure comprises at the macro-level only the components Solution and Results. So, deleting Results leads to a discourse structure in which only Solution is apparent. However, since Problem is implicit in the Solution discourse segment, this structure is the representation of the Problem-Solution sequence, as it is in the resulting summary produced by the application of heuristic 2. The above situation is not generalisable, i.e. the application of the heuristic 4 will not always lead to results similar to those provided by heuristic 2.

> **Hypothesis 5:** Justifications and evaluations are related to personal statements. In scientific discourse, evaluative statemens may be avoided for a very technical and objective communication.
>
> **Heuristic 5**: Delete Y, in X *Evaluation(X,Y)*

> **Hypothesis 6:** When a Solution unit does not correspond to a macro-component of the discourse (i.e. it appears in an intermediate level of the discourse structure), it emphasises the result of a particular solution for an inner problem (which is often implicit). So, the Result component is more significant than the Solution component[8].
>
> **Heuristic 6**: Delete Solution, in Solution-Result, when it is an intermediate discourse segment.

Applying the above two heuristics to the example summary, the result is

[8]Notice that this is the inverse case of Solution as macro-component: at the macro-level, (Problem,Solution) constitutes the minimum possible significant pair.

```
The interaction of continuous CO2 laser radiation focused onto a
free water surface is studied, both in normal gravity and in
reduced gravity conditions.  The depth of the keyhole structures
produced by different laser powers was observed.  The shape of
the keyhole and variation of depth with gravity are also
calculated.  There is a trend towards smaller bubbles at higher
power.  The speed of the momentum reaction flow down the side of
the keyhole has been calculated to be about 20 cm/s.  Very large
bubbles have been observed during a transition into low-gravity
conditions.
```

(96 words)

Other important heuristics include:

Heuristic 7: Delete Y, in *Justification(X,Y)*.
Heuristic 8: Delete Detail, in a Preview-Detail relation.
Heuristic 9: Delete Example, in a General-Example relation.
Heuristic 10: Delete X, in *Background(X,Y)*.
Heuristic 11: Delete Ca in a Ca-Co-P relation, when Purpose (P) is a repetition of Cause (Ca).
Heuristic 12: Delete X, in *Enables(X,Y)*.

A clearly careful analysis of the coverage of the outlined heuristics and hypotheses is required for integration into discourse strategies for generation of draft scientific summaries. Some of the aspects that deserve prompt consideration are:

- The recognition of essential discourse units in the discourse structure. In our analysis, we usually assume the leftmost unit of a relation is the most essential one. However, in some cases this assumption does not hold, e.g., in the Solution-Result/Eval relations in Figure 2, Solution is optional, according to Hypothesis 6.

- If a discourse component Y, in *Relation(X,Y)*, is a candidate for omission, Y can also be a candidate for absorption in a different position in the discourse structure. For example, if the relation Evaluation-Basis for Evaluation holds, and Basis for Evaluation is a list of Results (which is used to justify the evaluation), Basis for Evaluation could be displaced to a new discourse segment. This move would generate another discourse structure, and probably would correspond to another rhetorical effect.

- At an intermediate level of discourse representation, Matching relations correspond to contrasting relations and relations that introduce parallelism and compatibility between content units. It is unclear whether they may be omitted or not, due to the following factors:

    1. Matching Compatibility seems to introduce a kind of List relation (Scott and Souza, 1990), in which all the related discourse segments have equal weight. A specified heuristic must either prevent the suppression of any of these

components, or assume that any of them can be deleted at random. In this case, the deletion must be accompanied by linguistics markings of incomplete list, e.g. by using the verb include (a kind of relaxation of constraints takes place in this case, with no prejudice for meaning).

2. Concerning Matching Contrast, the omission of a contrasting statement is usually crucial, because meaning can be altered. If the same message is to be preserved, contrasting information should not be deleted.

We have not yet analysed Matching Parallelism relations.

Discussion

In this work, we follow a long tradition in discourse processing, which has been strongly influenced by work of Winter and Hoey. In computational linguistics, important work in this line has been carried out by describing a text structure as a tree of relations holding between pairs of spans of text (Hobbs, 1985; Mann and Thompson, 1987). Schema-based approaches have also oriented discourse organisation through combinations of rhetorical predicates (for example, (McKeown, 1985; Paris, 1993)).

In our work, we integrate different types of discourse relations in order to organise discourse coherently at a macro-level. We consider schemata appropriate for a precise account of coherence, and thus, for the production of suitable and understandable summaries[9]. Different schemata can correspond to different discourse strategies, so that different purposes of communication can be formulated and expressed to compress/expand discourse structures. Coherence can, thus, be enforced by the correct nesting and filling of schemata (Hovy, 1990).

The data provided by the linguistic analysis show that an average of sixteen relations hold in the corpus. From these, an average of 40% correspond to General-Particular or Matching relations; others equally important for compression are Enablement, Exemplification, Background, Evaluation, Justification, and Elaboration.

The "trial" set of heuristics for omission of information, based on the above relations, seems quite promising. The resulting summaries seem satisfactory, they are more indicative and significantly more condensed than the originals, and still convey coherent messages, by selecting content that is considered relevant in the context. Meaning is preserved, to the extent that it is possible to consider preservation of meaning in indicative summaries.

The typical communicative goals that seem related to compression refer mostly to *inform, indicate, justify* or *motivate*. For purposes of clarity, the specification of a metalanguage (as suggested by Hoey in his projection of segments of discourse into a sequence of questions and answers) seems also valuable. Combined with a universal representation of content, it can dominate choices of specific entities for any informative text and function as a powerful cohesive device, by allowing the introduction of markers of discourse relations (Winter, 1992). So, explicitness can also be addressed. Moreover, being defined according to the target community, the metalanguage also contributes to the observation of

[9]It is worth noting that, although schema-based approaches seem insufficient as a discourse model for certain types of discourse (see, for example, (Moore and Paris, 1993)), they are not problematic for the generation of summaries.

the readers' profiles, and thus, to their comprehension (Lehnert, 1981; Turk and Kirkman, 1989).

In this work, we deal only with structural features. For a complete generation framework, other heuristics that lead to concise, but still informative summaries must be specified, for a right balance between conciseness, informativeness and clarity. We are about to incorporate the heuristics presented in this paper into discourse strategies for the generation of draft scientific summaries. Lexicalisation and grammaticalisation of the discourse structures will be investigated in a later stage.

References

Endres-Niggemeyer, B. A Naturalistic Model of Abstracting. In Dagstuhl Seminar Report (9350), *Workshop on Summarising Text for Intelligent Communication*. Dagstuhl, Germany, 1993.

Francis, H. and Liddy, E.D. Structured Representation of Theoretical Abstracts: Implications for User Interface Design. In M. Dillon (ed.), *Interfaces for Information Retrieval and Online Systems: The State of the Art*. Greenwood Press, 1991.

Hobbs, J.R. *On the Coherence and Structure of Discourse*. Technical Report CSLI-85-37, Center for the Study of Language and Information, Stanford University, 1985.

Hoey, M. *Signalling in Discourse*. Discourse Analysis Monographs, English Language Research. University of Birmingham, 1979.

Hoey, M. *On the Surface of Discourse*. George Allen & Unwin (Publishers) Ltd., 1983.

Hovy, E. Unresolved Issues in Paragraph Planning. In R.Dale; C.Mellish and M. Zock (eds.), *Current Research in Natural Language Generation*, pp. 17-45. London, Academic Press, 1990.

Hutchins, J. *On the Structure of Scientific Discourse*. UEA Papers in Linguistics 5, pp. 18-39. University of East Anglia, 1977.

Hutchins, J. Summarization: Some Problems and Methods. In K.P. Jones (ed.), *Meaning: the frontier of informatics*, pp.151-173. Aslib, London, 1987.

Kintsch, W. and van Dijk, T. Toward a Model of Text Comprehension and Production. Psychological Review Vol. 85, No. 5, pp. 363-394, September 1978.

Lehnert, W.G. Plot Units and Narrative Summarization. *Cognitive Science*, 4, pp. 293-331, 1981.

Liddy, E.D. The Discourse-Level Structure of Empirical Abstracts: An Exploratory Study. *Information Processing & Management* 27(1), pp. 55-81, 1991.

Liddy, E., McVearry, K.A., Paik, W. Yu, E. and McKenna, M. Development, Implementation and Testing of a Discourse Model for Newspaper Texts. In *ARPA Human Language Technology Workshop Proceedings*. N.J., 1993.

Mann,W.C. and Thompson,S.A. *Rhetorical Structure Theory: A Theory of Text Organization*. Technical Report ISI/RS-87-190. California, June 1987.

McKeown, K.R. *Text Generation: Using Discourse Strategies and Focus Constraints to Generate Natural Language Text.* Cambridge University Press, 1985.

McKeown, K.R. Generating the Complex Sentences of Summaries Using Syntactic and Lexical Constraints: Two Applications. In Dagstuhl Seminar Report (9350), Workshop on *Summarising Text for Intelligent Communication*. Dagstuhl, Germany, 1993.

Moore, J.D. *A Reactive Approach to Explanation in Expert and Advice-Giving Systems*. PhD. Dissertation, University of California, 1989.

Moore, J.D. and Paris, C. Planning Text for Advisory Dialogues: Capturing Intentional and Rhetorical Information. *Computational Linguistics*, Vol.19, No. 4, pp. 651-694, 1993.

Paice, C.D. Constructing Literature Abstracts by Computer: Techniques and Prospects. *Information Processing & Management* 26(1), pp. 171-186, 1990.

Paris, C. L. *User Modelling in Text Generation.* Pinter Publishers, 1993.

Scott, D.R. and Souza, C.S. Getting the Message Across in RST-Based Text Generation. In R.Dale; C.Mellish and M. Zock (eds.), *Current Research in Natural Language Generation*, pp. 47-73. London, Academic Press, 1990.

Sparck Jones, K. *Discourse Modelling for Automatic Summarising*. Tech. Rep. No. 290. University of Cambridge, Feb 1993.

Sparck Jones, K. What might be in a summary? In Knorz, Krause and Womser-Hacker (eds.), *Information Retrieval 93*, pp. 9-26. Universitatsverlag Konstanz, June 1993.

Stoddard, S. *Text and Texture: Patterns of Cohesion.* Ablex Publishing Corp., 1991.

Turk, C. and Kirkman, J. *Effective Writing: Improving Scientific, Technical and Business Communication*. E. & F.N. Spon, New York, 1989.

Winter, E.O. The Notion of Unspecific versus Specific as one Way of Analysing the Information of a Fund-Raising Letter. In W.C. Mann and S.A. Thompson (eds.), *Discourse Description: Diverse Linguistic Analyses of a Fund-Raising Text*. John Benjamins Publishing Co., 1992.