COSY-MATS: An Intelligent and Scalable Summarisation Shell

Maria Aretoulaki
Dept of Pattern Recognition (Computer Science 5)
University of Erlangen-Nuremberg
Martensstrasse 3
D - 91058 Erlangen, Germany
aretoula@informatik.uni-erlangen.de

Abstract

In this paper, an architecture is presented for robust and portable summarisation, COSY-MATS COSY-MATS can avoid the superficiality and domain-dependence of IE approaches by means of high-level (pragmatic and rhetorical) content selection features It can also obviate the text type-dependence and cumbersome computation involved in NLU-based summarisation systems, because surface criteria are additionally used in the content selection process, as are identified mappings between those and the high-level features In this way, COSY-MATS should retain its generic and scalable character, while also permitting intelligent application-specific processing

1 Motivations behind the Design of COSY-MATS

The goal of the research reported here has been to develop a flexible, easily-portable and scalable, but also efficient and robust, NLP system that automatically generates summaries of real-world unrestricted texts To this effect, an architecture was designed for a hybrid **CO**nnectionist–**SY**mbolic **MA**chine for Text Summarisation (henceforth, **COSY-MATS**) (Aretoulaki, 1996)

A major concern in designing COSY-MATS has been to identify content selection features that are generic and application-independent (Section 2) The features should be applicable to any text, irrespective of domain or text type This is so that COSY-MATS is readily portable to different operation environments with a minimum amount of customisation The isolation of such features would provide a permanent infrastructure for both content selection and analysis The front-end text analysis modules can be developed so that they are geared towards the summarisation task, rather than text understanding in general, which is computationally-intensive Thus, these modules need only perform an analysis that is sufficient for the evaluation of the selected content selection features The establishment of universal importance determination criteria means that the permanent set of analysis, interpretation, content selection and generation processors can be extended with application-specific modules during the porting of COSY-MATS This is also what renders COSY-MATS a type of *summarisation shell* Significantly, the computations of the supplementary modules will already be accommodated for in the standard flow of processing of the system by virtue of these features (Section 3)

Admittedly, the identification of content selection features of general applicability is a very difficult task This is demonstrated in the limitations of the two main trends in current summarisation research (cf (Aretoulaki, 1996)) There are *Information Extraction* (IE) environments, which perform a superficial and partial analysis of the input text based on the progression of keywords and application-specific phrasal patterns therein, e g (BT, 1994, Jacobs and Rau, 1990, Luhn, 1958, MUC-5, 1993, Paice, 1981, Paice, 1990, Salton et al , 1994) The problem with IE systems is that, although they can be used very efficiently on any type of text, they are domain-dependent and likely to produce inaccurate output This is due to their excessive reliance on specialised content words There are also systems which are based on *Natural Language Understanding* (NLU) methods involving deeper processing Apart from syntactic and lexico-semantic analysis, the hierarchical rhetorical organisation of the source text can also be taken into account, as can certain aspects of the context of the discourse, e g (Garighano et al , 1993, Lehnert, 1981, Mitkov et al , 1994) Such more sophisticated types of system, however, are prohibitively slow as a result of the extensive processing involved They are also very fragile, because the high-level knowledge employed is usually hand-coded and hence arbitrary and incomplete Even when this knowledge has been acquired automatically, e g (Maybury, 1993, Soderland and Lehnert, 1994), it is application-dependent Consequently, despite their occasional

domain-independence (e g (Endres-Niggemeyer and Neugebauer, 1995, Ono et al , 1994, Rau et al , 1993, Sharp, 1991)), NLU approaches are —on the whole— specialised in a particular text-type

For the design of COSY-MATS, a *holistic* and unifying approach has been adopted that involves both extralinguistic, NLU-type, analysis and selective statistics-based linguistic processing reminiscent of IE, in co-ordination Similarly to NLU, analysis in COSY-MATS is sufficiently deep for the semantic, rhetorical and contextual aspects of the input text to be considered in content selection In contrast to what the case is with such systems, however, the computation of these diverse aspects of the text is efficient This is because objective cues on the surface of the text are also exploited in COSY-MATS, echoing the IE approach Nevertheless, unlike IE, these cues are function words and generic content words which point towards the high-level functions of the respective textual units in the context of the discourse, while at the same time being domain-independent Thus, apart from identifying universal content selection criteria that should render COSY-MATS portable and scalable, the research reported here has also attempted to establish mappings between the concrete and the more abstract criteria in the devised feature scheme, so that the system is also intelligent and practical, i e so that the evaluation of these abstract criteria is fully automated (Section 2)

2 Intelligent Content Selection Criteria

In order to identify generic content selection features that can be used by COSY-MATS in any application context, an extensive corpus analysis was carried out on a variety of real-world texts Three main types of text have been analysed *newspaper articles, scientific papers* and *(scientific) author abstracts* The subcorpus of newspaper articles (160) is extremely diverse in both its content and form The topics range from business news and legal reports to social commentary, medical issues and politics Similarly, the other two subcorpora consist of 170 articles and abstracts, respectively, that pertain to scientific fields such as computer science, the natural sciences, as well as philosophy and linguistics In addition, the texts are of varying length from half a page in the case of the abstracts and most news agency reports, to four or more pages, when scientific papers and newspaper special reports are involved Consequently, apart from covering a range of subject domains, the corpus used in designing the content selection processes in COSY-MATS also represents more than two text types

The corpus was analysed both on the surface and on more abstract levels Given the diversity of the types of text and the writing styles exhibited in the corpus, regularities regarding the rhetorical development of the texts and the central informational units therein could not be easily established Only in the case of the scientific papers and their abstracts could any statements be made on the logical progression of the presentation of the content, from the purpose of the research, to the methodology, the experimental set-up and the evaluation of any results (cf (Gopnik, 1972, Jordan, 1993, Lucas et al , 1993, Maizell et al , 1971)) The newspaper articles were mainly studied in terms of groups of adjacent sentences and the rhetorical relationship between them (cf (Ono et al , 1994)) No generalisations could be made regarding their top-level organisation

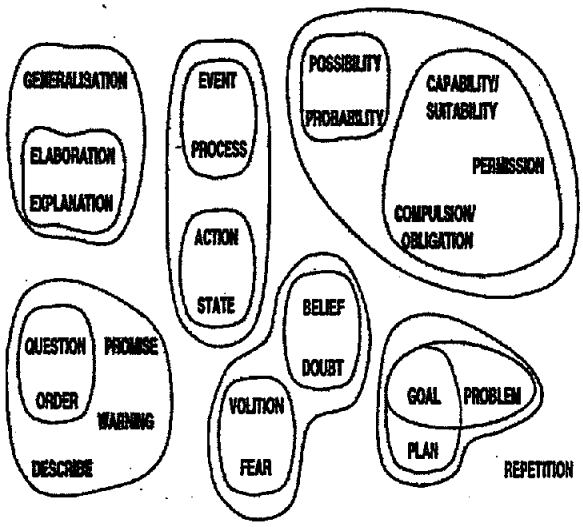
A number of theories of pragmatics, discourse analysis and text development have provided useful concepts for this study of the corpus at a higher level

- a) theories which are preoccupied with *the communicating agents*, their goals, plans and beliefs, such as Speech Act Theory (Austin, 1962, Searle, 1969), Rhetorical Structure Theory (RST) (Mann and Thompson, 1987), or AI research on scripts (Lehnert, 1981, Schank and Abelson, 1977) and belief ascription (Wilks and Ballim, 1987)
- b) theories on the tracking of the *discourse history* by means of identifying the focused items therein, e g (Grosz, 1986, Hobbs, 1978, Reichman, 1985, Sidner, 1983, Webber, 1983)
- c) theories of *cohesion and coherence* and how these are manifested on the surface of the text, e g Systemic-Functional Linguistics (Halliday and Hasan, 1976) and the Problem-Solution information metastructure (Hoey, 1994, Jordan, 1984) (cf (Paice, 1981))

The diversity of the subject matter covered in the corpus has meant that specialised keywords were ignored in its analysis Instead the emphasis was placed on function words and regular general-language content words which are associated with the instantiation of the semantic, rhetorical and pragmatic functions considered Such lexical items can be employed as markers, not only of the development of the discourse but also of the focused and central points therein In this process, the various cohesion and coherence theoretical frameworks were very influential, as were the computational approaches to focus prediction and identification

As a result of this corpus analysis at the surface and more abstract levels, 87 features have been identified as relevant to content selection and importance determination across domains and, largely, text types (Aretoulaki, 1996) Three descriptive levels are used for their classification the *pragmatic*, the *intermediary* and the *surface*, in decreasing order of abstraction The three levels reflect, in a sense, the three main trends in discourse theory identified Thus, the 24 pragmatic features

(Fig 1) encode information related to the communicating agents Pragmatic features such as *Plan* and *Goal*, for instance, are reminiscent of AI work on scripts (Schank and Abelson, 1977), *Elaboration* and *Explanation* can be parallelled to RST relations (Mann and Thompson, 1987)



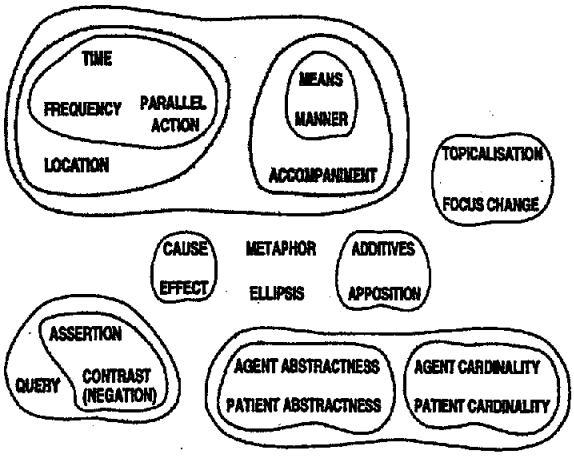
Clusters of Pragmatic Features

Figure 1

The intermediary features (Fig 2) represent rhetorical semantic criteria often employed in the processing of focus information and in anaphor disambiguation For example, *Topicalisation, Focus Change, Cardinality* and *Ellipsis* have all been used in computational contexts such as (Hobbs, 1978, Reichman, 1985, Sidner, 1983, Webber, 1983)

Finally, the surface features (Fig 3) coincide mostly with explicit cues in the text which denote cohesive and coherence relations among sentences (cf (Luhn, 1958, Paice, 1981)) The *Function Word* and the *Common Content Word Pools*, for instance, consist of lexical items with a semantic/rhetorical load extensively discussed in a Systemic-Functional (Coulthard, 1994) and Problem-Solution context (Jordan, 1984, Jordan, 1995) Consequently, by using features such as these in COSY-MATS, all three levels of language —from the low-level surface to the high-level pragmatic— can be collectively considered in order to 'holistically' determine the importance of individual propositions in a text

Apart from this grouping of the features into different levels, the surface and the intermediary features proposed in this scheme have also been used to objectify the abstract pragmatic features This was in order to facilitate the automatic evaluation of the latter during the actual operation of the fully-



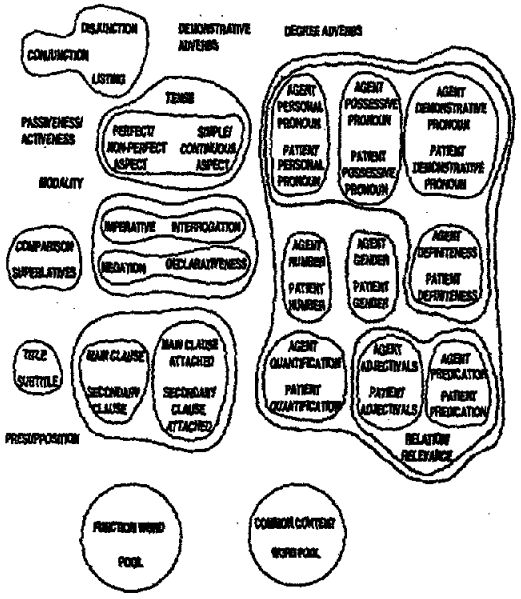
Clusters of Intermediary Features

Figure 2

developed COSY-MATS (cf Section 3) To this effect, a number of interlevel mappings were identified both between the pragmatic and the lower levels, and between the intermediary and the surface levels These mappings were compiled in a manual which was used by 5 subjects in encoding texts from this corpus (Aretoulaki, 1996) The encoded texts were then employed for the empirical testing of a prototype of the content selection module, reported in Section 4 Example mappings are given below

- The pragmatic feature *Repetition* is correlated to the surface features *Personal* and *Possessive Pronouns* and *Demonstratives* (Sidner, 1983) It is also associated with the intermediary *Focus Change* (Sidner, 1983, Webber, 1983) and *Ellipsis* (Hovy, 1987) This is because the central topics in a text are often resumed by means of anaphora, both in the same sentence and later on in other important sentences
- The presence of impersonal phrases in the *Passive* on the surface level is extensively used to express a *Generalisation* on the pragmatic level The latter denotes a central text unit by definition (Gopnik, 1972, Lehnert, 1981, van Dijk and Kintsch, 1978)
- The surface *Negation* is correlated to the intermediary *Contrast* (Jordan, 1984)
- *Modals* such as "should" are also extensively used on the surface of discourse, when proposing, evaluating, or making tentative claims in general Thus, this feature is also related to the pragmatic *Belief/Doubt, Volition/Fear* and *Plan* (cf (Fukumoto and Tsuji, 1994))

Evidence for the usefulness of the interlevel mappings proposed in the context of the COSY-MATS con-



Clusters of Surface Features

Figure 3

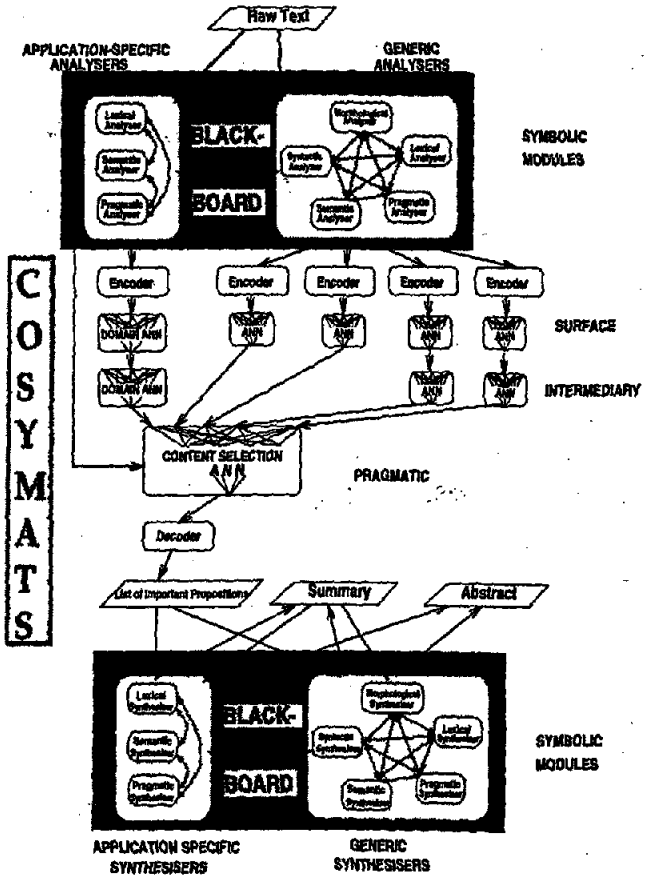


Figure 4 The Architecture of COSY-MATS

tent selection feature scheme was provided by validation tests regarding the uniformity of the feature evaluation practices among the human encoders (Aretoulaki, 1996) The encoding of an identical part of the corpus by means of all the pragmatic features showed that there was a total of 79 6% agreement among the encoders on the evaluation of the pragmatic features, using the above-mentioned manual Consequently, the identified surface and other less subjective features can be fully exploited later on for the automation of the encoding of the abstract pragmatic features The validation tests also indicated that there was 96% agreement on which of the corpus sentences were important and which unimportant for the corresponding texts

3 A Scalable Architecture for Intelligent Summarisation

Having identified 'universal' content selection features, as well as some of the ways these interact with each other, the following architecture was designed for a full-scale implementation of the COSY-MATS summarisation shell (Fig 4) (Aretoulaki, 1996) Every sentence in the text to be summarised[1] is first processed by a cluster of standard symbolic analysers, morphological, syntactic, semantic and pragmatic The result of this processing is the evaluation of a set of basic linguistic and extralinguistic features that provide the input for a cascade of low and higher-level Artificial Neural Networks (ANNS), each responsible for specific subtasks The low-level ANNs map linguistic features (surface and intermediary) into extralinguistic features (intermediary and pragmatic) The pragmatic features provide the input to the highest-level content selection ANN that ultimately determines the relative degree of importance of each sentence This latter ANN is also the only component of COSY-MATS that has been implemented to date Finally, the sentences selected as important during the content selection phase will be used as the basis for generating either a comprehensive summary or a more concise abstract (Aretoulaki, 1996) This processing will take place in another cluster of symbolic processors, almost symmetric to that used for text analysis and interpretation It is here that the planning and the actual synthesis of the summary/abstract will be realised However, it is important to note that the output list of the best-scoring sentences produced by the content selection ANN can also be used to provide a *draft summary*, i e a concatenation of already-existing sentences instead of an original text (cf (Kupiec et

[1] which is assumed to be integral and coherent, rather than a random collection of propositions,

al , 1995)) This is also the only type of generation that is currently preoccupying this research (cf Section 4 1)

Despite the dominance of the generic modules therein, COSY-MATS does provide for the incorporation of application-specific information First of all, the architecture is highly modular, so that new specialised processors can be —in principle— simply plugged in The simplicity of the interface between the various modules means that new modules that are either symbolic or connectionist can equally well be accommodated For example, in addition to the existing lower-level ANNs, other ANNs can be easily incorporated which have been trained to recognise specific keywords and structural phrases that differentiate one domain or text type from the other in expressing the same rhetorical and pragmatic functions Hence, COSY-MATS can function as a shell for the building of specialised summarisers

As regards the front-end symbolic analysers, the processing that will take place therein will be dictated by the type of data that needs to be computed in the ANNs The latter computation, in turn, will be based on the identified generic and application-specific mappings across the three levels of description the pragmatic, the intermediary and the surface (Section 2) In addition, it is the implementation of the content selection ANN that will determine the eventual type and number of pragmatic features required for the whole process of summarisation (Section 4) As a result, a partial analysis and interpretation of the input text only need to be performed in COSY-MATS The common problem in NLU-based systems of combinatorial explosion and inefficient computation in the search for a solution will thus be largely avoided At the same time, this pragmatism in the analysis and interpretation processes does not decrease the amount of deep processing (semantic, discourse and pragmatic) carried out in the system High-level processing is salient in the pragmatic features identified These are, nonetheless, 'grounded' by means of the generic lower-level features, as well as other surface and semantic characteristics of texts pertaining to the specific application of interest

In summary, the proposed architecture is both modular and hybrid The complex task of content selection is systematically decomposed into much more manageable computations In addition, the strong points of both symbolic and connectionist processing are combined in a complementary way (cf (Aretoulaki, 1996)) The symbolic analysers can work with structured data of arbitrary length laden with variables They also have powerful symbol-matching facilities (as is appropriate for lower-level text analysis) In contrast, the ANNs are able to deal with fuzzy and inexact processing (as is involved in importance determination and inter-level feature mappings) (McClelland and Rumelhart, 1986, Rumelhart and McClelland, 1986)

4 Empirical Evidence

As the first and most crucial step in implementing COSY-MATS, a prototype of its content selection ANN was developed This is a standard feed-forward back-propagation network (Rumelhart et al , 1986) This ANN receives individual text sentences from the text to be summarised, hand-coded[2] by means of the identified pragmatic features, and assigns to them degrees of importance It has been a major assumption behind this work that it is *feature combinations* rather than individual features that characterise sentence importance (Sections 1 & 2) An ANN learns such interactions naturally, which is why the connectionist paradigm was adopted for the content selection task

The training corpus consisted of 1,880 sentences in total, taken from the real-world text collection described in Section 2 1,100 of them are sentences largely out of their context, while the remaining 780 sentences make up 29 full texts In contrast to the diversity of the former subcorpus, each of the latter texts is approximately 23 sentences long and was fully encoded The encoding was carried out by 5 individuals on the basis of the above-mentioned manual which exemplifies the correlations between the surface and the more abstract features in the proposed scheme The manual was used in order to standardise the encoding process as much as possible, as well as to validate the proposed ways in which the evaluation of the abstract pragmatic features can be objectified and fully automated later on in the completed system

Experiments to date (cf (Aretoulaki, 1996)) have demonstrated the superiority of the pragmatic features over input to the ANN from across the three levels of abstraction (*58 1%* vs *56 1%* success on average, where 'success' coincides with agreement with the judgement made by the human encoder regarding the level of importance of the corresponding sentence) The simultaneous use of control experiments with noisy data[3] has ensured the validity of these results (50 1% success) In addition, the testing on whole texts has provided comparable results to those acquired with isolated sentences, namely *56 8%* success on average, this suggests that the pragmatic features are sufficiently abstract to capture hierarchical and structural aspects of the corresponding discourses

The diversity of the corpus in terms of subject matter, text type and length provides sufficient evidence for the appropriateness of the pragmatic fea-

[2] given that the remaining components of COSY-MATS have not been implemented as yet,

[3] These used characteristics of the text that should be irrelevant to the content selection task, such as 'The second word in the sentence ends in a vowel'

tures for the high-level representation of texts from any domain or text type Moreover, the portability of these pragmatic content selection features has also been partly proved with experiments on whole texts (Aretoulaki, 1996) These indicated that only a small amount of retraining is required for the ANN to deal with new text types, which involves a limited number of representative texts Thus, what is predicted to differ between text types is the relative influence of each of the identified features in the final weighting of the corresponding sentence

4.1 Generating Draft Summaries

The 'draft' summaries that result after concatenating the sentences of the input text that were selected by the ANN as important are, on the whole, adequate for current awareness purposes (See (Aretoulaki, to appear) for a detailed evaluation of this and other draft output) The ANN receives a single —coherent and largely cohesive— text each time, rather than a collection of unrelated texts Sentence selection was based on the 24 pragmatic features used for their encoding and the statistical correlations among them, as indicated in the training corpus Most importantly, by filtering out the sentences for which the ANN did not have a clear decision, i e by adapting the corresponding threshold on-line, content selection can be more fine-grained and the output summaries more brief An example draft summary for a newspaper article after the application of this type of filtering is shown below In this case, there was *82 6%* agreement between the ANN decision and the corresponding human judgement regarding the importance of individual sentences in this article[4]

> **(1)** Moscow editors feel the old-fashioned grip of the state (Headline)
>
> **(3)** Intense party pressure for the dismissal of a prominent liberal editor and a new campaign to discredit the radical politician Boris Yeltsin - both apparently with the backing of President Gorbachev - have raised fears among reformers of a conservative swing by the Soviet leadership **(5)** On Monday evening, he was summoned to the Central Committee to be told in so many words by Vadim Medvedev, the Politburo member in charge of ideology, that he should leave his post **(6)** The move follows a harsh talk delivered last week by Mr Gorbachev to a group of senior Soviet editors, in which he gave several a dressing down **(12)** Some journalists are talking of a protest strike **(13)** 'The press is quite simply now facing bans on what it can write about, we're going back to the situation of years ago,' one complained yesterday **(16)** The motion, which could prefigure a head-on clash between the party and a steadily more assertive parliament, attacks the Central Committee ideology department for its 'unacceptable attempts' to cow a newspaper **(22)** Backing for Mr Yeltsin is not universal **(23)** But the fact that the parliamentary exchanges were broadcast on prime time television leaves no doubt that a campaign is under way to smear a man whose huge following makes him Mr Gorbachev's only real rival

Despite the coincidental cohesiveness therein, this draft output comprises the majority of the semantically substantial sentences in the input text The concatenation of sentences from the original is undoubtedly a much simpler task than the generation of an extended summary or a concise abstract Novel text synthesis in the fully-developed COSY-MATS will also benefit from the proposed mappings between the surface and the more abstract content selection features Since the corresponding modules, however, have not been implemented yet, the processes involved will not be exemplified here

5 Conclusion: COSY-MATS is not a Utopia

All experimental results to date indicate that content selection in the completed COSY-MATS environment can be robust and efficient, even in the absence of any customisation to the specific application (domain or text type) or the user requirements This is due to the adoption of the connectionist paradigm for this fuzzy task and the proven generic nature of the pragmatic and lower-level features used therein

In the context of further implementing this summarisation shell, current work includes the testing of alternative learning algorithms for the prototype content selection ANN in order to improve its success rate In addition, the more rigourous specification of the mappings between the surface cues and the intermediary and pragmatic features is attempted for the subsequent development of specialised processors that compute them Thus, the encoding of the pragmatic features will be fully automated and it will also be possible to measure the precise effect that this will have on the training of the whole cascade of ANNs, given the current practice of hand-coding Moreover, the impact on the content selection ANN of incorporating application-dependent information in the system will also be studied (cf (Aretoulaki, 1996)) What is important is that research to date has proved that the realisation of the COSY-MATS intelligent and scalable summarisation shell is by no means a utopia

[4]The 5 subjects were free as to the number of sentences they could pick out from any text as important Importance, in turn, was defined as the relative indispensability from the final summary of the propositions expressed in the corresponding sentence This was determined on the basis of the whole text the sentence belongs to

6 Acknowledgements

The research reported in this paper was carried out as part of a Ph D programme at the Centre for Computational Linguistics at U M I S T I am indebted to my supervisor, Prof Jun-ichi Tsujii, for his invaluable feedback and encouragement during that time I am also grateful to CANON Europe Ltd for granting me with a 2-year Research Studentship, without which this research would not have ever been possible Finally, I would like to thank the two anonymous reviewers for constructive comments which have increased the degree of clarity of this paper

References

M Aretoulaki *COSY-MATS A Hybrid Connectionist - Symbolic Approach To The Pragmatic Analysis Of Texts For Their Automatic Summarisation* Ph D Thesis, Dept of Language Engineering, U M I S T, Manchester, U K, March, 1996

M Aretoulaki *A Hybrid Connectionist-Symbolic Approach To Pragmatics and Text Summarisation* University College London (UCL), London, U K, To Appear

J L Austin *How To Do Things With Words* Oxford University Press, Oxford, U K, 1962

Science and Technology Feature Short Cuts *The Economist*, pages 97–98, 1994 December 17th

M Coulthard, editor *Advances in Written Text Analysis* Routledge, London, U K, 1994

B Endres-Niggemeyer and E Neugebauer Professional Summarising No Cognitive Simulation without Observation In *Proceedings of the 4th International Colloquium on Cognitive Science (ICCS-95)*, Donostia, San Sebastián, Spain, 1995

J Fukumoto and J Tsujii Breaking Down Rhetorical Relations for the Purpose of Analysing Discourse Structures In *Proceedings of the 15th International Conference on Computational Linguistics (COLING-94)*, volume 2, pages 1177–1183, Kyoto, Japan, August 1994 Association for Computational Linguistics

R Garigliano, R G Morgan, and M H Smith The LOLITA System as a Contents Scanning Tool In *Proceedings of the 13th International Conference on Artificial Intelligence, Expert Systems and Natural Language Processing*, Avignon, France, May 1993

M Gopnik *Linguistic Structures in Scientific Texts* Mouton, The Hague, the Netherlands, 1972

B J Grosz The Representation and Use of Focus in a System for Understanding Dialogs In B J Grosz, K Sparck Jones, and B L Webber, editors, *Readings In Natural Language Processing* Morgan Kaufmann, California, 1986 First appeared in Proceedings of the 5th IJCAI, Cambridge, Massachusetts, 67-76, Los Altos William Kaufmann, 1977

M A K Halliday and R Hasan *Cohesion in English* Longman, London, U K, 1976

J R Hobbs Resolving Pronoun References *Lingua*, 44 311–338, 1978

M Hoey Signalling in Discourse A Functional Analysis of a common Discourse pattern in written and spoken English In M Coulthard, editor, *Advances in Written Text Analysis*, pages 26–45 Routledge, London, U K, 1994

E Hovy Generating Natural Language Under Pragmatic Constraints *Journal of Pragmatics*, 11 689–719, 1987

P S Jacobs and L F Rau SCISOR Extracting Information from On-Line News *Communications of the ACM*, 33(11), pages 88–97, 1990

M P Jordan *The Rhetoric of Everyday English Texts* George Allen & Unwin, London, U K, 1984

M P Jordan Openings in Very Formal Technical Texts *Technostyle*, 11(1) 1–28, 1993

M P Jordan The Power of Negation Pragmatics, Discourse Patterns and Clausal Semantics In *Annual Meeting of the Canadian Association of Teachers of Technical Writing*, Canadian Learned Conferences, Montreal, Canada, 1995

J Kupiec, J Pedersen, and F Chen A Trainable Document Summarizer In *Proceedings of the 18th ACM-SIGIR Conference*, pages 68–73, 1995

W G Lehnert Plot Units and Narrative Summarization *Cognitive Science*, 4, 1981, pages 293–331

N Lucas, K Nishina, T Akiba, and K G Suresh Discourse Analysis of Scientific Textbooks in Japanese A Tool for Producing Automatic Summaries Technical Report 93TR-0004, Dept of Computer Science, Tokyo Institute of Technology, Tokyo, Japan, March 1993

H P Luhn The Automatic Creation of Literature Abstracts IBM Journal of Research & Development, 2 (2), April, 1958, pages 159–165

R E Maizell, J F Smith, and T E R Singer *Abstracting Scientific and Technical Literature An Introductory Guide and Text for Scientists, Abstractors, and Management* Wiley-Interscience John Wiley & Sons, Inc, New York, 1971

W C Mann and S A Thompson Rhetorical Structure Theory A Theory of Text Organization Technical Report, ISI Reprint Series ISI/RS-87-190, USC Information Sciences Institute, Marina Del Rey, Ca, June 1987

M T Maybury Automated Event Summarization Techniques In Dagstuhl-Seminar-Report 79 *Summarising Text for Intelligent Communication*, B Endres-Niggemeyer, J Hobbs, and K Sparck-Jones, editors, Dagstuhl, Wadern, Germany, Dec 13–17, 1993

J L McClelland, D E Rumelhart, and the PDP Research Group (Eds) *Parallel Distributed Processing Explorations in the Microstructure of Cognition Volume 2 Psychological and Biological Models* The MIT Press, Cambridge, Massachusetts, 1986

R Mitkov, D Le Roux, and J -P Desclès Knowledge-Based Automatic Abstracting Experiments in the Sublanguage of Elementary Geometry In C Martin-Vide, editor, *Current Issues in Mathematical Linguistics*, pages 415–421 North-Holland, the Netherlands, 1994

Proceedings of the Fifth Message Understanding Conference (MUC-5), Aug 25-27, San Mateo, CA, Baltimore, Maryland, 1993 Morgan Kaufmann

K Ono, K Sumita, and S Miike Abstract Generation based on Rhetorical Structure Extraction In *Proceedings of the 15th International Conference on Computational Linguistics (COLING-94)*, volume 1, pages 344–348, Kyoto, Japan, August 1994 Association for Computational Linguistics Also appeared as cmp-lg/9411023

C D Paice Automatic Generation of Literature Abstracts - An Approach Based on the Identification of Self-Indicating Phrases In R N Oddy, S E Robertson, C J van Rijsbergen, and P W Williams, editors, *Information Retrieval Research*, pages 172-191 Butterworths, London, U K , 1981

C D Paice Constructing Literature Abstracts by Computer Techniques and Prospects In *Information Processing & Management*, 26(1), pages 171–186, 1990

L F Rau, R Brandow, and K Mitze Domain-Independent Summarization of News In Dagstuhl Seminar Report 79 *Summarising Text for Intelligent Communication*, B Endres-Niggemeyer, J Hobbs, and K Sparck-Jones, editors, Dagstuhl, Wadern, Germany, Dec 13–17, 1993

R Reichman *Getting Computers to Talk Like You and Me Discourse Context, Focus, and Semantics (An ATN Model)* The MIT Press, Cambridge, Massachusetts, 1985

D E Rumelhart, G E Hinton, and R J Williams Ch 8 Learning Internal Representations by Error Propagation In D E Rumelhart, J L McClelland, and the PDP Research Group, editors, *Parallel Distributed Processing Explorations in the Microstructure of Cognition Volume 1 Foundations*, pages 318–362 The MIT Press, Cambridge, Massachusetts, 1986

D E Rumelhart, J L McClelland, and the PDP Research Group (Eds) *Parallel Distributed Processing Explorations in the Microstructure of Cognition Volume 1 Foundations* The MIT Press, Cambridge, Massachusetts, 1986

G Salton, J Allan, C Buckley, and A Singhal Automatic Analysis, Theme Generation, and Summarization of Machine-Readable Texts *Science*, 264 1421–1426, 1994 3 June

R C Schank and R P Abelson *Scripts, Plans, Goals, and Understanding An Inquiry into Human Knowledge Structures* Erlbaum, Hillsdale, NJ, 1977

J R Searle *Speech Acts An Essay in the Philosophy of Language* Cambridge University Press, Cambridge, 1969

B Sharp INFORMEX - An Information Extractor System In *Proceedings of the International Conference on Current Issues in Computational Linguistics*, pages 361–371, Penang, Malaysia, June 1991

C L Sidner Focusing in the Comprehension of Definite Anaphora In M Brady and R Berwick, editors, *Computational Models of Discourse*, pages 267 – 330 The MIT Press, Cambridge, Massachusetts, 1983

S Soderland and W Lehnert Corpus-Driven Knowledge Acquisition for Discourse Analysis In *Proceedings of the 12th National Conference on Artificial Intelligence (AAAI-94)*, Vol 1, pages 827–832, Seattle, Washington, Jul 31 - Aug 4, 1994

T A van Dijk and W Kintsch Cognitive Psychology and Discourse Recalling and Summarizing Stories In W U Dressler, editor, *Current Trends in Textlinguistics*, volume 2 of *Research in Text Theory*, pages 61–80 de Gruyter, Berlin, 1978

B L Webber So, What Can We Talk About Now? In M Brady and R Berwick, editors, *Computational Models of Discourse*, pages 331–371 The MIT Press, Cambridge, Massachusetts, 1983

Y Wilks and A Ballim Multiple Agents and the Heuristic Ascription of Belief In *Proceedings of the 10th International Joint Conference on Artificial Intelligence (IJCAI-87)*, pages 119–124, Milan, Italy, August 1987