Brigitte Endres-Niggemeyer, Jerry Hobbs,
Karen Sparck Jones (editors):

Summarizing Text for Intelligent Communication

Dagstuhl-Seminar-Report; 79 [Full version]
13.12.-17.12.93 (9350)

ISSN 0940-1121

Tel.: +49 - 6871 - 2458
Fax: +49 - 6871 - 5942

Das Internationale Begegnungs- und Forschungszentrum für Informatik (IBFI) ist eine gemeinnützige GmbH. Sie veranstaltet regelmäßig wissenschaftliche Seminare, welche nach Antrag der Tagungsleiter und Begutachtung durch das wissenschaftliche Direktorium mit persönlich eingeladenen Gästen durchgeführt werden.

Bezugsadresse: Geschäftsstelle Schloss Dagstuhl
Universität des Saarlandes
Postfach 15 11 50
D-66041 Saarbrücken, Germany
Tel.: +49 -681 - 302 - 4396
Fax: +49 -681 - 302 - 4397
e-mail: office@dag.uni-sb.de
www: http://www.dag.uni-sb.de/dagstuhl.html

B. Endres-Niggemeyer, J. Hobbs, K. Sparck Jones (editors):

Summarizing Text for Intelligent Communication

Dagstuhl Seminar 13.12.-17.12.1993, Dagstuhl, Germany

Dagstuhl-Seminar-Report; 79

Most of the participants

Due to cost reasons, this seminar report exists in three versions:

- Printed Short Version (Dagstuhl-Seminar-Report; 79)
- Printed Full Version for participants
- Full Electronic Material

Table of Contents

of the long (printed) version of the Dagstuhl seminar report about

Summarizing Text for Intelligent Communication

Availability of the Seminar's Material

Note:

Due to cost reasons, this report has been divided up into a short version (with a higher number of copies), a quite lenghty volume for the participants, and an electronic version containing all material.

Printed versions can be ordered from the

Dagstuhl office
Universität des Saarlandes
Postfach 15 11 50
66041 Saarbrücken, Germany.
office@dag.uni-sb.de
+49-681-302-4396 phone
+49-681-302-4397 fax

The materials of Dagstuhl seminar 9350 are available

- in **printed** form (from Dagstuhl office):

  Short Report
  (Rationale, Program, Question List, Outcome, Participants)

- on the **Internet**
  (Full Version) via WWW and FTP from the following sites:

  Dagstuhl office (main site)
  http://???.dag.uni-sb.de/???/Abstract/
  ftp://???/pub/???/Abstract.tar.gz
  (URL to be announced, mail to office@dag.uni-sb.de for current details and other mirror sites)
  Polytechnic of Hannover, Information and Communication Department (home site)
  http://www.bid.fh-hannover.de/SimSum/Abstract/
  ftp://www.bid.fh-hannover.de/pub/Abstract/Abstract.tar.gz
  University of the Saarland, Philosophical Faculty, Information Science (mirror)
  http://www.phil.uni-sb.de/FR/Infowiss/Abstract/
  ftp://ftp.phil.uni-sb.de/pub/Abstract.tar.gz

The organizers gratefully acknowledge

- the generous support of the seminar by the Deutsche Forschungsgemeinschaft (German Science Foundation)
- the contributions to the publication of this report made by the University of Cambridge and the Polytechnic of Hannover
- the untiring effort of Alexander Sigel, the seminar's technical assistant, who also put our report on paper and on the web.
- the Dagstuhl team's friendly and flexible commitment to all matters of the seminar and its participants.

The Seminar Rationale

Summarising is a vital information processing task. We have proposed a Dagstuhl seminar on automatic summarising, now, for the following reasons:

- there has been a rapid growth of interest and activity in discourse and text processing in general and summarising in particular
- there has been useful progress in computational text interpretation and generation, for example through the Message Understanding Conferences
- there is a chance to integrate recent and pertinent results from related disciplines, from cognitive science to document retrieval, and to rethink system design in the light of their findings
- there is a manifest need to broaden the approaches used in existing summarising systems, and to adopt new ones, to obtain systems able to handle texts that vary in length, topic type, and form
- there is a clear requirement, for intelligent communication in man–machine interaction, for summaries tailored to specific user needs.

More powerful summarising systems than those developed so far are clearly needed. Today's information systems are not able to summarise in an intelligent way, deriving significant information from their text sources in order to provide the user with material of an appropriate scale, depth and orientation, in a coherent text form. The few systems built so far have been limited in both approach and implementation. They have been based either on shallow, statistically–oriented approaches to identifying salient source content, or on deeper analysis but only within a prespecified topic framework. They have thus been either only weak and uncertain in their ability to capture key information, or effective only within a very narrow application context and not readily extensible.

But summarising is not just a primary task for text handling systems. It is also called for as a subtask in many other information management contexts, for instance in interactive consultation or instruction, so building a flexible summarising capability into systems for these purposes will enhance their performance. At the same time, since summarising depends on discourse interpretation, transformation and generation, it is a crucial test of discourse theory. Developing adequate theories of discourse structure and processing for summarising will thus benefit all areas of language processing.

The Dagstuhl seminar we report on is built on the premise that summarising is a complex,

knowledge–based task to which many different language, text, and world knowledge resources jointly apply. This view of summarising as a composite task implies an interdisciplinary seminar, bringing together both those who are directly engaged with summarising and those contributing to parts or aspects of the whole.

We therefore invited researchers working on automatic summarising, practitioners with experience of abstracting and information retrieval, researchers engaged with text retrieval, computational linguists working on methods of text analysis and production, members of the NLP community attacking specific tasks like message processing and data extractions, or concerned with user interfaces, linguists studying pragmatics and discourse, and psychologists and cognitive scientists especially concerned with text and discourse processing. All of these have an important contribution to make, both to the development of an intermediate summarising technology operating on text, and to longer term research on summarising for which full text understanding is required.

Organisation of the Seminar

Given that we see summarising as a composite task, we organised our specialised international and interdisciplinary workshop in sessions devoted to these five major themes:

1. Human Summarising (HS) – practice and models:

- Empirical research on abstracting
- Summarising strategies
- Textual knowledge processing

2. Automatic Summarising (AS) – implementation and systems:

- Automatic summarising systems
- Text and message processing
- Text typology

3. Related Disciplines (RD) – discourse analysis and use:

- Discourse theories
- Text typology
- Text processing and knowledge use
- Location of information etc.

4. User Adaptation (UA) – needs and strategies

5. Computational Resources (CR) – tools and processes:

- Text and message processing
- Text representation

Final Seminar Program
(13.12.93)

Summarizing Text for Intelligent Communication

Building a research platform for theoretical and practical progress in summarizing, as a key task in natural language processing, artificial intelligence, and related disciplines

Date: Dec. 13-17, 1993

Organizers:
Karen Sparck Jones, Univ. of Cambridge (chairperson)
Brigitte Endres-Niggemeyer, Polytechnic of Hannover (organizer)
Jerry Hobbs, SRI International, Menlo Park
Elizabeth Liddy, Syracuse University
Cecile Paris, ISI Marina del Rey

Overview

- INTRODUCTORY SESSION
- HUMAN SUMMARIZING – practice and models: I, II, III
- AUTOMATIC SUMMARIZING – implementation and systems: I, II, III
- RELATED DISCIPLINES – discourse analysis and use: I, II, III, IV
- COMPUTATIONAL RESOURCES: I, II
- USER ADAPTION – needs and strategies
- EVALUATION METHODS for Summarization
- A RESEARCH PLATFORM FOR INTELLIGENT SUMMARIZING: I, II

Monday 13-12-1993

9.00 - 10.30
INTRODUCTORY SESSION
Chairperson: Wolfgang Wahlster
Rapporteurs: Brigitte Endres-Niggemeyer / Paul Jones

Karen Sparck Jones: Introduction *(using a paper by Hutchins)*
Karen Sparck Jones: Summarising: analytic framework, key component, experimental method

11.00 - 12.30

HUMAN SUMMARIZING – practice and models I

Chairperson: Gerhard Strube
Rapporteur: Sumiko Mushakoji
Hans Strohner: Inferences in Text Processing: Summaries and Instructions
Rosemarie Gläser: Summarizing Texts as Genres of Academic Writing

14.00 - 15.30

HUMAN SUMMARIZING – practice and models II

Chairperson: Hans Strohner
Rapporteur: Ines Busch-Lauer
Helmut Felix Friedrich: Training of Reductive Text Learning Strategies
Edward Cremmins: Valuable and Meaningful Text Summarization in Thoughts, Words, and Deeds

16.00 - 17.30

HUMAN SUMMARIZING – practice and models III

Chairperson: Elizabeth Liddy
Rapporteur: Helmut Felix Friedrich
Harold Borko and Brigitte Endres-Niggemeyer: An Empirical Process Model of Abstracting

19.00 Guitar Concert and Informal Welcome Party

Tuesday 14–12–1993

9.00 - 10.30

AUTOMATIC SUMMARIZING – implementation and systems I

Chairperson: William Black
Rapporteur: Harold Borko
Udo Hahn: Concept-Oriented Summarizing in the Text Condensation System TOPIC: 12 Claims and 6 Desiderata for Design
Ellen Riloff: A Corpus-Based Approach to Domain-Specific Text Summarization: A Proposal

11.00 - 12.30

AUTOMATIC SUMMARIZING – implementation and systems II

Chairperson: Udo Hahn
Rapporteur: Ralph Weischedel
Lisa F. Rau: Domain-Independent Summarization of News
Woojin Paik: Chronological Information Extraction System (CIES)

14.00 - 15.30

AUTOMATIC SUMMARIZING – implementation and systems III

Chairperson: Patrizia Violi
Rapporteur: Annely Rothkegel
Elizabeth D. Liddy: Development and Implementation of a Discourse Model for Newspaper Texts
Mark T. Maybury: Automated Event Summarization Techniques

16.00 - 17.30

RELATED DISCIPLINES – discourse analysis and use I

Chairperson: Raya Fidel
Rapporteur: Ines Busch-Lauer
Jerry Hobbs: Summaries from Structure

Wednesday 15–12–1993

9.00 - 10.30

RELATED DISCIPLINES – discourse analysis and use II

Chairperson: Nicholas Belkin
Rapporteur: Wojin Paik
Annely Rothkegel: Abstracting in the Perspective of Producing a Text
Livia Polanyi: Linguistic Dimensions of Text Summarization

11.00 - 12.30

RELATED DISCIPLINES – discourse analysis and use III

Chairperson: Livia Polanyi
Rapporteur: Ellen Riloff
Bruce Britton: Summarizing Situation Models. Using Principal Components to Reconstitute the Expert's Causal Model in the Reader's Mind

14.00 - 15.30 EXCURSION to DFKI, Saarbrücken: System Demonstrations

16.00 - 17.30

18.00 Conference Dinner

Thursday 16-12-1993

9.00 - 10.30

RELATED DISCIPLINES – discourse analysis and use IV

Chairperson: Donia Scott
Rapporteur: Lucia Rino
Sumiko Mushakoji: Constructing 'Identity' and 'Differences' in Different Scientific Texts and their Summaries: Its Problems and Solutions
Ines Busch-Lauer: Abstracts in German Medical Journals - a Linguistic Analysis

11.00 - 12.30

COMPUTATIONAL RESOURCES I

Chairperson: Cecile Paris
Rapporteur: Mark Maybury
Kathleen R. McKeown: Generating the Complex Sentences of Summaries Using Syntactic and Lexical Constraints: Two Applications
William Black: Parsing, Linguistic Resources and Semantic Analysis for Abstracting and Categorisation
[John A. Bateman: Using text structure and text planning to guide text summarization] *(paper was not presented)*

14.00 - 15.30

USER ADAPTATION – needs and strategies

Chairperson: Wolfgang Wahlster
Rapporteur: Edward Cremmins
Raya Fidel: User-Centered Text Analysis
Nick Belkin: On the Relationship between Discourse Structure and User Intention

16.00 - 17.30

COMPUTATIONAL RESOURCES II

Chairperson: Kathleen McKeown
Rapporteur: Elisabeth Maier
Ralph Weischedel: From Text to Objects to Summaries

20.00 - 21.30

EVALUATION METHODS for Summarization

(Group Discussion)

Friday 17-12-1993

9.00 - 10.30:

A RESEARCH PLATFORM FOR INTELLIGENT SUMMARIZING I

Chairperson: Jerry Hobbs

Rapporteur: Brigitte Endres-Niggemeyer
The Cognitive View of Discourse and Text:
A Pragmatic Strategy for Summarizing Systems

11.00 - 12.30:

A RESEARCH PLATFORM FOR INTELLIGENT SUMMARIZING II

Chairperson: Karen Sparck Jones

Rapporteur: Brigitte Endres-Niggemeyer
The Research Agenda for Intelligent Summarizing

End of the seminar

Rapports of the Sesssions and Extended Abstracts

(in Session Order)

Introductory Session

Rapporteurs: Brigitte Endres–Niggemeyer and Paul Jones

Wolfgang Wahlster:

Introduction to the Seminar

In text summarization, both human and automated, a variety of research has been conducted. However, nothing approaching a science of summarization has been achieved. A major problem for the field is the absence of any simple quality standards. For instance, it is not known how an optimal summary may be defined. In comparison with text summarization as a whole, the subfield of information extraction seems more advanced.

It seems necessary to expand the scope of summarization which is still often restricted to the processing of written text. E.g., summarization may be oral or multimodal, it may use a complete or partial representation of the original information, and we may summarize tables, databases, images or events just as well as written texts.

The task of the seminar is to contribute to the development of summarization research and therefore help the research community move towards better scientific methods and achievements.

John Hutchins:

Text Summarization

(presented by Karen Sparck Jones)

Summarizing can be described according to four sets of parameters:

- coverage (individual documents or text collections)
- informativeness (indicative, informative, evaluative summaries)
- selectivity (abstracts reporting the original completely or in part)
- recipients (for special user group, for general consumption)

With varying emphasis, the following subprocesses determine summarization:

- selection of what is "important"

- ommission of what is "unimportant"
- generalization from the particular to the specific
- identification of general (global) structures)

Karen Sparck Jones:

Summarizing: Analytic Framework, Key Component, Experimental Method

A summary is a condensed derivative text, i.e. a content reduction by selection or generalisation of what is important in the source.

Summaries are conditioned by their input, purpose, and output factors. The input factors refer to a range of form and subject–type properties, the purpose factors to the summary's audience and function, the output factors to material of the source considered for summarizing (complete, or selected), and format (continous or itemized).

The basic architecture of a summarizing system comprises two main components:

1. The source text is interpreted into a source representation. The first involves individual sentence analysis and then a global representation of the input text.
2. Summary generation. Here one forms a summary representation from the source representation, and then synthesizes the summary text from it.

Along with source content, large–scale structure plays an important part in summarizing.

Short "noddy" texts were represented using three different representations: RST (Mann & Thompson), "predication participation" (Lehnert), and Intentional Structure (Grosz & Sidner). For summarizing, the way proposed by the representation led to alternative summaries of the 10 short test texts. All test summaries had some plausabilities, so they may constitute a good starting point for further research.

Introduction to "Text Summarization" workshop

John Hutchins

1. Why summaries?

Why are summaries made at all? The simple answer is that in order to gain access to and control the flood of information, everyone needs to know briefly what is worth reading, what is useful for a particular purpose. Nobody wants to waste time reading what is useless. In giving an overview of content, summaries save time. It is not, we must stress, a new phenomena – it began even before the invention of printing. Scholars in the early Middle Ages found it necessary to compile summaries of what was known. These were the first encyclopedias. With this long history, it is somewhat amazing that the process of summarization has itself been so neglected by scholars. Only relatively recently, with access to computers capable of dealing with large textual databases, has there been serious research in this field. This is the theme of this workshop.

As an introduction I have attempted to 'summarize' the basic features.

2. A typology of summaries can be made on four sets of parameters:

(a) coverage
(b) informativeness
(c) selectivity
(c) recipients

1. Text coverage.

Summaries can be made of individual texts (documents, speeches, presentations, books) or of collections of texts (proceedings of a debate, arguments of prosecution and defense in a law court, documents in an archive, collections of papers or books, etc.)

2. Informativeness.

On this parameter we may identify three types: indicative, informative, and evaluative. An informative summary reports on the factual data conveyed or the details of the opinions expressed. An indicative summary states only that certain

topics were covered without conveying the content of the facts or opinions described. An evaluative summary locates the content or opinions of the text within the context of other texts treating similar topics.

3. Selectivity.

Summaries can attempt to cover all 'important' aspects of texts (global) or they can be selective (reporting only part of the content). Typically, selective summaries are made for specific purposes or clientele.

4. Recipients.

Summaries can be made for specific groups of recipients or readers (directed summaries), i.e. they can be targeted to their particular needs – by being evaluative or selective summaries. They are possible only when the specific needs of users are known or predictable. However, summaries can also be for general consumption (undirected), i.e. for use in an information system where the background knowledge of users cannot be predicted. In such cases, summaries are normally indicative. The production of informative summaries intended for 'unpredictable' needs is probably the most difficult of all.

5. In theory each of the **parameters may combine**. In practice, some parameters are found more commonly together than others, e.g. directed summaries are usually evaluative or selective; general summaries are normally indicative or informative.

3. Methodology of summarization.

In the broadest terms, summarization can be regarded as a conjunction (with varying emphasis) of the following processes:

(a) selection of what is 'important'
(b) omission of what is 'unimportant'
(c) generalization from the particular and specific
(d) identification of general (global) structures

Methodologically, summaries can build up from the details to generalizations (by generalization, selection, and omission) or can work from global frameworks. In evaluative and selective summaries the first two processes dominate. But all informative summaries involve the last two. The major difference comes with the indicative summaries. These

may not necessarily generalise at all.

Automatic summarization was initially based entirely on the first two methods: extraction of what it is hoped is 'important' (on the basis of textual cues or triggers) and thus the omission of everything not extracted. This method is still widely used alone or in conjunction with other methods

More sophisticated and more recent attempts at automatic summarization involve 'generalization', i.e. attempts to derive general statements of content from particular textual elements using (typically) semantic/lexical hierarchies, networks, thesauri and macrostructural frameworks, and (in many cases) databases of domain knowledge. Usually, the subject ranges of such experiments are highly restricted, and apparently not easily transferable to other domains.

Attempts to automate indicative summarization are rare. Whereas informative summaries are based on methods which consider the whole text as source data for deriving summaries (whatever the methods used), indicative summaries would be based on those parts of texts which state the 'topics' (i.e. what the text/sentence, etc. is 'about'). They would ignore those parts which convey what is 'new' ('rheme'). In some early efforts at automatic summarization, the 'topics' of texts were selected from the initial sentences of paragraphs. Although the aim was probably the production of 'informative' summaries, in effect the method was more appropriate to indicative summarization. A more sophisticated approach to indicative summarization would involve the identification of thematic progression in texts (i.e. the ways in which sentences relate to each other in terms of thematic and rhematic links) and the identification of 'theme' sentences in paragraphs and in texts as wholes.

Purposes/roles of summarization

Some contexts in which summarization is important (and the type(s) which are appropriate):

1. Abstracts

Abstracts are a vital component of communication of research, saving the reading time of individual researchers and improving the control of information, e.g. in computer–based 'free text' document retrieval systems.

They provide in addition important clues and sources of index terms (keywords) in bibliographic databases. (In practice, most automatic indexing systems are based

not on the full texts of documents but on their abstracts – produced manually, of course.)

Abstracts can be informative or indicative. In general, abstracts produced by authors themselves are invariably informative, but also selective. Abstracts produced by others can be either informative or indicative. If intended for a general–purpose database (e.g. bibliography or abstract journal) they are normally non–evaluative and non–selective. If intended for a specific audience, they will normally be both selective and evaluative.

Indicative abstracts are most appropriate in circumstances where the background of users is unknown. I have argued elsewhere that indicative abstracts based on the 'thematic' approach (outlined very briefly above) can provide the best 'starting points' for users who are seeking documents in unfamiliar fields of knowledge. They know what the 'themes' should be but not what the 'rhemes' are going to be. (An informative abstract is appropriate for the researcher looking for documents treating subjects at the forefront of research – mentioned in the 'rhemes' of texts.)

2. Review articles

Typically an article reviewing progress in a specific research field covers a wide range of documents. In some cases, only the bare contents of texts are mentioned (indicative); in others, more substance is reported (informative). But most importantly, the review article weighs up the current status and indicates the important contributions (evaluative and selective).

3. Encyclopedias

Encyclopedia articles review the state of the art in a more global fashion. They are evaluative (and almost necessarily selective) summaries of 'what is known' about a particular topic (informative). They represent 'starting points' for readers, and hence frequently refer to other encyclopedia articles or 'further reading' (in this respect they are indicative).

4. Legal summaries

The trial summaries by lawyers and judges are examples of evaluative summaries of informative nature. The summaries of prosecuting and defendant lawyers are necessarily and deliberately evaluative (and normally selective); the summaries of judges when speaking to juries are supposedly neutral, non–evaluative, but equally selective.

5. Journalism

Summaries are the stock in trade of most journalists. Many newspaper reports are extracts from other texts (e.g. reports of debates and official documents). Most journalist summaries are selective (often evaluative).

6. Market surveys/reports

These basic information sources for business people are intrinsically compilations of (evaluative) summaries of documentation produced by companies and of evaluations of products.

7. Translation

While the translating of a text is not as such a summary, it can be argued that what the recipients of translations often need is not a translation of the whole text but something which conveys the basic message (essence) of the text in the unknown language. In other words, they do not want just a translation but also a summary. The conjunction of machine translation and automatic summarization would be an obvious and valued desideratum for future NLP research. (Whether MT itself offers methods applicable to Automatic summarization per se is another question.) In addition, the demand for multilingual access to databases is growing rapidly: MT can offer help in searching databases in unknown languages as well as (obviously) translating the results of searches.

5. In view of the central function of summarization in society it is **surprising that it is a topic which has been so often neglected**. (Indeed, it could be argued that summarization is central to all communication: every expression of a 'theme' is in essence a 'summary' of some preceding statement – in either the same 'text' or in some other earlier 'text'.)

It is a personal pleasure to me that this workshop is taking place, since it is devoted to a topic in which I have had an interest for over twenty years (although without being actively involved in research as such). The workshop is bringing together researchers from a wide range of disciplines, and I am sure that discussions in the coming week will bear fruit in both the immediate and more distant future.

References

Hutchins, W.J. (1970)

Linguistic processes in the indexing and retrieval of documents. *Linguistics 61*, 1970, 29–64

Hutchins, W.J. (1977a)

On the problem of 'aboutness' in document analysis. *Journal of Informatics 1(1)*, April 1977, 17–35

Hutchins, W.J. (1977b)

On the structure of scientific texts. *UEA Papers in Linguistics 5*, September 1977, 18–39.

Hutchins, W.J. (1985)

Information retrieval and text analysis. In: T.A.van Dijk (ed) *Discourse and communication* (Berlin: de Gruyter, 1985); 106–125.

Summarising: analytic framework, key component, experimental method

Karen Sparck Jones

Computer Laboratory, University of Cambridge
New Museums Site, Pembroke Street, Cambridge CB2 3QG, UK
sparckjones@cl.cam.ac.uk

November 1993; SERC Grant GR/F 26478

What is a summary? How is summarising done? We make and use summaries all the time, but automatic summarising has not made much progress. Systems so far have been either highly application specific or crudely approximative. My aim in recent work has been lay some foundations for less restricted and more powerful systems, first by providing a framework for investigating the core summarising process, and second by experimenting with alternative approaches to the core process. The framework analysis lays out the context factors affecting summarising and outlines a basic architecture for studying the core process. The experiments with the core concentrate on text structure, and specifically on large–scale text structure, as a crucial contributor to summarising.

1 Analytic framework

Work on automatic summarising so far has often relied on assumptions that have not been examined, and research on summarising and in relevant areas like discourse has been very heterogeneous. The analytical framework I have begun to develop is intended on the one hand to make the factors affecting summarising explicit, so it is easier to identify those that are especially important and to consider how they may be automated; and on the other hand to make it possible to assess and compare, and a way that is productive for automation, different approaches to summarising and the characterisation of texts that these require.

I shall present this as a framework for *text* summarising both for convenience and as more immediately relevant to automation. But it is intended to be a general framework, as further considered later. Thus as a starting point we can say that a summary text is a condensed derivative of a source text, reducing content by selection and/or generalisation on what is important. We want to explicate this definition of a summary and model the process of obtaining it.

1.1 Context factors

Summaries are conditioned by input, purpose, and output factors.

The *input* factors, characterising the source text, cover a range of *form* and *subject*–type properties. Form properties include structure, scale, medium, and genre; subject properties are categorisable as 'ordinary', 'specialised', or 'local', independent of specific topic domain. (Source readership, if explicit or inferable, could also be an input factor.) Even with a common notion of summary, and summarising process, summarising will be influenced by these global properties of the source text as well as by its individual content: compare, for instance, those of a university medical textbook with those of a magazine gardening column.

The *purpose* factors in summarising refer to a summary's *audience* and *function*. The former may be definable as 'targetted' or 'untargetted', the latter refer to alternative types of objective for a summary, notably to impart or to alert. Thus compare a novel summary in a newspaper with a company report digest for investment analysts.

Finally, the *output* factors subsume *material*, that is whether the summary range over the source is deliberately comprehensive or partial, and structure, scale, medium, and genre; the summary range over the source is deliberately comprehensive or partial, and *format*, namely whether the summary itself is continuous or itemised. Compare a biological paper summary giving test results with an archaeological report summary under standard headings.

This framework is only a rough scaffolding: thus many factor values, like genre, are only indicative. But the framework makes the point that general descriptive properties of summaries in relation to their sources, e.g. whether they are reflective or reorganising, stem from both the character of the source and the objective of the summary. However even with better factor specification, it does not follow that particular combinations of input and purpose factor values determine output ones: output choices have also to be made, not least because summaries are linguistic objects with their own requirements to meet. More generally, factor specification cannot be made tight enough, even when taken in conjunction with source content, to determine the individual summary. This applies whether summaries are constructed in advance, for possible future consumption, or on the spot to suit an individual current need.

Thus the main point of the framework is on the one hand to emphasise the fact that summarising is context dependent and that the source–summary relationship is not autonomous; and on the other the fact that the context constraints can only be broad ones and that there is no one correct summary of a source text. There are many legitimate distinct (and not just trivially different) summaries of the same source text, even for some type of context defined by a particular factor combination and, in the limit, for some individual context. The implication for summary *evaluation* is

that this must refer to *utility*, over unique occasions of use; so saying that a summary correctly captures its source content is either a prediction about likely utility in many situations or a remark about observed utility in some specific situation: even when a summary is made for a unique occasion, it is only as good as it happens to be on that occasion.

1.2 Architecture

At the same time, because the context constrains the summary, it is necessary to see how a context specification is applied to the summarising process. This requires an account of processing itself. The other part of the framework is therefore a view of the essential architecture of a summarising system, primarily grounded in the transformation of source text to summary text as an operation on text content.

In this, content refers to both *information* and *expression*, and summarising as a general notion may be distinguished from other types of source reduction by being equally concerned with both, implying a need for both *coherence* and *cohesion* in the summary. A precis places more emphasis on expression than information, while abstracting often does the reverse.

The basic architecture of a summarising system has two major components, each with two subcomponents. The source text is *interpreted* into a *source representation*, involving first individual sentence analysis and then integrating analyses to form a global representation for the input text. Summary *generation* also has two stages, forming a *summary representation* from the source one, and synthesising the output summary text. The formation step, which may invoke information not in the source itself, is clearly the crucial one.

This is a logical account: it does not imply, for instance, that analysis for the whole text is completed before integration can begin. The essential point is that the source text representation is independent of the summary representation. This separation is in general required, even when source interpretation is done only for the summarising task, since the summary representation cannot be constructed until the complete source is available. Moreover, though having a complete source representation to work on is appropriate for automated summarising, the general requirement for an independent source representation is compatible both with the position that the source representation itself embodies a summary (cf Kintsch and van Dyck), and that a summary may be dynamically constructed and reconstructed, during discourse processing.

2 Key component

Though the foregoing may appear obvious and uncontroversial, the framework is useful both as a means of clarifying the assumptions made in individual cases (for example DeJong's use of only a

summary representation), and of investigating the different elements of summarising. In particular, it allows us to address the central summarising questions, designed to decompose our initial characterisation of a summary, in an organised and hopefully productive way. These questions are:

1. what data does a source representation make available for summarising; and
2. how are these data used for summarising? (or, what does a summary require of a source representation and how is this supplied?)

More specifically, what general properties of a source text are there, as opposed to its specific content, that are relevant to summarising and are independent both of input factor values and of (at least all but the most eccentric) summarising purposes?

Answering this question implies experiments that maximise comparisons across different input factor combinations as well as across differences of individual document content, and minimise the constraints of summary purpose. Then, given some resulting insight into key properties of text and natural ways of using these in summarising, we can investigate the role of context factors and the derivation of tailored summaries.

2.1 Large–scale structure

Given the definition of summarising as a language–processing task, and that it is manifest there is *large–scale structure* in discourse, it seems clear that such structure has an important part, along with specific source content, to play in summarising. My recent work, with John Beaven, has been devoted to studies of different kinds of large–scale structure, the kinds of source representation they provide, and the sorts of ways these can be used for summarising, as follows.

Broadly speaking, we have three *types* of source text information, *linguistic*, *domain*, and *communicative*, and hence may have large–scale structure of each type. For example, we can see the same text in terms of linguistic organisation and categories as, say, a comparison with one illustration followed by another; in terms of its discourse domain structure as providing property characterisations of objects; or in terms of its communicative intent as motivating a preference for one purchase rather than another. We also have alternative representation *forms*, *top down* and *bottom up*, defined here as instantiated versus constructed and not just in processing terms. For instance an object characterisation may be an instance of a generic frame, or simply an adhoc predicate aggregation. We thus have a *grid* of six kinds of source representation.

3 Experimental method

To begin to compare these alternatives for source representation, and assess their value for

summarising, we have taken a (modest) set of test texts of different sorts and carried out experiments as follows.

To obtain source representations we have worked from a common *baseline* representation consisting of sentence representations in a (simplified) conventional logical form (as e.g. delivered by the SRI Cambridge Core Language Engine), with anaphors resolved in the Sidner style. Thus the baseline representation for a source text is simply a predication network. We have then obtained elaborated *full* representations, capturing large–scale structure as defined by *exemplars* for the type–form pairs drawn as far as possible from the discourse analysis and summarising literature. Thus we assigned exemplars to slots in the type–form grid and used all of these for each test text, taking the texts' input factor properties as uninspected givens.

For the bottom–up cases we used Mann and Thompson's Rhetorical Structure Theory, 'predication participation' (cf Lehnert), and Grosz and Sidner's Intentional Structure respectively as linguistic, domain, and communicative exemplars. For the top–down cases we used Yale–type scripts and frames as domain exemplar; but we could not simply apply rhetorical schemas as used for generation to interpretation, and therefore examined a middle–level story grammar approach following Rumelhart instead. We have not yet found a top–down communicative exemplar to test. These grid assignments are crude, but are defensible; the exemplars are also not necessarily optimal, but are suitable candidates to start with. Applying the exemplars to obtain representations for the test texts has necessarily been done by simulation, because most of these approaches have never been computationally implemented, or would require significant resource provision for the test materials; but we were also more interested in the output representations than in how they might be obtained, though when providing scripts, for example, we deliberately sought generality.

For summarising, we initially adopted definitions for the summary representation that followed naturally from those of the large scale structure. For instance, taking the ultimate clause for the topmost nucleus for Rhetorical Structure Theory, or the representation for the dominant intention in an Intentional Structure. We thus adopted a 'default' approach to summary purpose. For the output text, we assumed a simple state of the art text generator.

We have therefore obtained a set of alternative summaries for each of the ten test texts and can consider how far they satisfy our intuitions about capturing what is important in these, and whether they capture the same or different source content: different *strategies* may give different results or, alternatively, important source content may be so clearly important that any strategy will pick it up. However we could not consider the structure of the output summary text as our inputs were too short to generate extended outputs. We have also not considered either the generic character or the specific detail of the summarising process itself: for instance whether summarising is a sort of cyclic process (as in human summarising), or how local focus information is handled, as well as

how the specific strategies would be automatically implemented.

3.1 Assessment

The test summaries all have *some* plausibility; however the test texts are very short. The individual strategies all showed both hits and misses, in representation and in summarising, and when compared both similarities and differences. The natural next steps are to extend these tests to cover more data and strategies; to compare performance between more elaborate approaches simple, shallow ones suited to practical implementation; and to see what combining strategies using distinct information types would involve and deliver; and to study factor influences.

Gladwin, P., Pulman, S. and Sparck Jones, K. 1991
Shallow processing and automatic summarising: a first study, *Technical Report 223*, Computer Laboratory, 1991.

Sparck Jones, K. 1993
'What might be in a summary?' in *Information Retrieval '93* ed Knorz, Krause and Womser–Hacker, Universitätsverlag Konstanz, 1993.

Human Summarizing – practice and models I

Rapporteur: Sumiko Mushakoji

Dr. Hans Strohner and Dr. Gert Rickheit

In the framework of cognitive systems, in which information is processed in animal–like manner. Dr. Hans Strohner and Dr. Gert Rickheit described some basic inference processes in humans and computers. Their focus is on inferences when processing simple texts, in particular, when trying to comprehend summaries.

They took instructions for special examples of summaries, and introduced six kinds of inferences; sensory–motor inferences, systactic inferences, code inferences, reference inferences, semantic–sense inferences, and pragmatic inferences.

The cognitive system moved us from studying the summerized products along the certain linguistic aspects to studying the very process of comprehending summaries as the multiple complex inference processes.

Dr. Rosmarie Gläser

Dr. Rosmarie Gläser introduced a powerful notion "genre", which comprises, in John Swales' statement, *"a class of communicative events, the member s of which share some set of communicative purposes"* (Swales, 1990). She focused on practical ways of teaching summarizing texts as genres in classroom.

She outlined a variety of the original text genres at first, and then defined an essential characteristic of summerizing text genres; they select, evaluate, order and condense items of information to their relevance for a particular subject or a particular purpose. From this point of view, the linguistic act of summarizing is a complex cognitive process which is composed of different cognitive operations. Dr. Gläser regards that these processes are ultimately reflected in a text.

Recognizing summarizing as a cognitive and linguistic process, Dr. Gläser proposed an outline of teaching summarizing genres in courses of Academic Writing. Her conclusion is that the choice of genres and sub–genres which have a summarizing function is open–ended, and will vary from term to term according the students' needs.

There were two major points in this session:

1. **How should we take the notion (definition) of "source of summaries"?**
   Dr. Strohner and Dr. Rickheit have extended the notion of the source of summary from text to the processes observed themselves. This notion was controvertial during the discussion. Dr. Gläser emphasized the importance of source text "genre".
2. **The importance of focusing on the process of summarizing (not the products) is stressed, but to which direction empirical studies are going on?**
   Dr. Strohner and Dr. Rickheit's experiment is under way. Dr. Gläser's teaching summerizing will produce many findings in the classroom.

Inferences in Text Processing: Summaries and Instructions

Hans Strohner and Gerd Rickheit

Sonderforschungsbereich 360
"Situierte Künstliche Kommunikatoren"
Universität Bielefeld

1. Introduction

Some shortcomings of traditional linguistic approaches can be avoided by analysing the language processing system in the framework of cognitive systems theory. A cognitive system is a concrete system which is able to process information in an animal–like manner. According to this definition, not only human beings, but also computers and robots that simulate the cognitive behaviour of humans belong to the class of cognitive systems.

The main dimensions of a cognitive system are its architecture, its dynamics, and its development. Let us look first at the architecture. If we want to analyse the architecture of a specific cognitive system, we must take into consideration its components, its environment, its structure and its function:

- A cognitive system has two main components, namely the information processor and the processed information. The main components of the information processor are cognitive models, i.e. cognitive units that represent circumscribed parts of the environment or of the system itself. The processed information consists of various information relations, i.e. properties of objects that point to other objects, as a footprint in the snow points to the animal that made it. In a cognitive system which is able to process linguistic information, the informational component of the system is constituted by a linguistic object, for example a text or an utterance.
- The environment of a cognitive system comprises all those objects that are functionally related to its components. This functional impact may be related to the object properties of matter, energy or information.
- Just as the components and the environment of a cognitive system are inseparably intertwined, so too are structure and function. Structure and function interact intimately and thus build up a relational network which constitutes the knowledge of the cognitive system. The structural part of the knowledge functionally intervenes between input and output. The behaviour of the system is thus largely influenced by the knowledge of the system as an important part of its overall structural and functional state.

On the basis of the general concept of information, it is possible to distinguish several specific types of information that may appear in a cognitive system:

- The sensory–motor information connects the processor with the object and its informational content.
- The syntactic information is a relation between the components of the object.
- The semantic information is a relation between the perceived object and those objects which have contributed to the object information.
- The pragmatic information is a relation between two or more information processors.

The information types are represented in the information processor as specific knowledge types, i.e. as sensory–motor, syntactic, semantic and pragmatic knowledge. Since the semantic knowledge relates the text and other objects in the environment by means of cognitive models in the information processor, it is a very complex relation consisting of the following three parts:

- the code relation, which connects the two cognitive models of the information and the information source internally;
- the reference relation, which connects the information source with its cognitive model;
- the sense relation, which connects this model with the other relevant world knowledge of the information processor.

Only when all three semantic relations are taken into account is a complete description of the semantic knowledge possible from a cognitive point of view. Non–cognitive semantic analyses tend to neglect at least one of the proposed semantic aspects.

The intensive research into the dynamics of a cognitive system during the last few years has demonstrated that there are many different types of processes. They include automatic and intentional processes, bottom–up and top–down processes, forward and backward processes, and excitation and inhibition processes. All of these interact in very complex ways partially influenced by the situation, context, the background knowledge, and the emotional state of the information processor. This complexity applies even more to the developing processes in a cognitive system.

In this paper we describe some basic inference processes in humans and computers when processing simple texts. In particular, we focus upon those inferences drawn when people summarize a text or when they try to comprehend summaries. From a certain perspective instructions can be taken for special examples of summaries. Since instructions generally have immediate consequences, researchers have a good chance of observing some of the processes going on when instructions are comprehended. In our research group a model of the processing of

instructions was developed, which is called ACTIO (Artificial Communication Theory of Instructions and Operations).

2. Sensory–motor Inferences

The sensory–motor knowledge links the language processor with the situated text by means of the auditory or visual input and motor output processors. This knowledge is crucial for the processor–environment interaction and provides the basis for the perceptual processing of the text.

In the perceptual processing of summaries and instructions, the printed layout is important. The lay–out determines what will catch the reader's eyes. In many cases, pictures are helpful not only to capture attention but also to present additional information. In the ACTIO model, the processor tries to combine all available sources of information in order to build up an integrated knowledge of the environment.

3. Syntactic Inferences

Only slowly has a theory of syntax processing been developed that has done away with the restrictions of syntactic structuralism. The cognitive systems approach assigns the syntactic structure an auxiliary function in processing semantic and pragmatic relations. Syntactic ambiguity problems, such as the attachment of prepositional phrases, are solved not only by syntactic processes but also by semantic strategies which take into consideration the relevant world knowledge.

4. Code Inferences

Inferences at the code level of language processing decide upon the right entrance into the semantic domain of text information. One main task of code inferences is to find out the context–specific meaning of ambiguous words. Another task is the construction of the internal architecture of the lexical concept.

As in summaries a certain degree of abstraction is needed, authors should take care to select lexical items which are concrete enough to give the reader clear information on the topic described.

5. Reference Inferences

Although code inferences are very important for the semantic processing of linguistic information, they are certainly not sufficient for semantic analysis. What has to be added is referential connection of linguistic information to the external world and the construction of mental models of

this world. The inferences by means of which these tasks might be achieved are the reference inferences.

Referential inferences are especially useful when the referential relations of the concepts in the texts are vague, as is often the case in summaries. In addition to the basic reference inferences we need special inferences for those concepts which are connected to each other by means of coreferential relations. Research has shown that coreference depends not only on syntactic relationships, but often on world knowledge, too.

6. Semantic Sense Inferences

Semantic sense inferences combine different semantic units. These units may be concepts, propositions, or larger units of discourse. With semantic sense inferences the processor is able to build a coherent representation of the text meaning even if the text itself offers only fragmentary information (Rickheit & Strohner, 1985, 1992). The importance of a statement can only be considered as a part of coherent text information.

A type of semantic sense inference which is highly relevant for the processing of summaries is the instantiation of category terms. Do readers generally infer shark when they read The fish attacked the swimmer? Whereas earlier studies with recall methods lead to the con– clusion that readers routinely infer instantiations from general concepts, more recent studies with more adequate methods are more sceptical. Instantiation inferences were found only when typical specimen served as targets and when these specimen are in the focus of the text information.

At the level of local coherence, sense inferences connect parts of, or whole, propositions. Several studies investigated whether people draw inferences on implied case–filling elements, such as agents, patients, and instruments, while reading. The question is, for example, whether the participation of a dentist is inferred when reading the sentence The tooth was drilled. It was shown that such case–filling inferences are drawn only very rarely.

Causal connections range among the most important relationships between events. Again, experimental results indicate that the consequences of actions are only inferred if the inference contributes to the coherence of the text.

The sense inferences which we have discussed so far are based on relatively simple relations between concepts or propositions. We have seen that inferences are drawn mainly in order to make the text more coherent. One may speculate that inferences will be more problable when they are based on more complex semantic structures, such as scripts, scenarios, or mental models. The relevant research, however, gives no definite answer to this question (McKoon & Ratcliff, 1992).

7. Pragmatic Inferences

Pragmatic inferences connect the text knowledge with the mental model of the partner of the language processor. Thus the intention of the text author may be recognized. In this important field of text processing conclusive experimental evidence is still lacking. One reason for this deficiency may be due to methodological problems of experimental research.

8. Conclusion

Inferences play an essential role in the processing of summaries. In this paper, we have discussed several types of such inferences, including sensory–motor, syntactic, code, reference, semantic sense, and pragmatic inferences. Most of these processes may be present in the reader of a summary, but they also should somehow be present in its author if she or he wants to make the statements comprehensible to the reader. The inferences discussed are relevant not only for summaries, but also for instructions which are in many respects comparable to summaries.

9. References

McKoon, G. & Ratcliff, R. (1992).
Inference during reading. *Psychological Review*, 99, 440–466.

Rickheit, G. & Strohner, H. (Eds.) (1985).
Inferences in text processing. Amsterdam: North Holland.

Rickheit, G. & Strohner, H. (1992).
Towards a cognitive theory of linguistic coherence. *Theoretical Linguistics*, 18, 209–237.

Summarizing Texts as Genres of Academic Writing

Rosemarie Gläser

Leipzig

The topic under discussion poses a number of problems and calls for a clarification of concepts and terms.

1. The linguistic status of genre

We define *genre* as a prototypical class of texts. An *LSP genre* (Fachtextsorte) may be defined as an historically established, institutionalized and productive pattern for the logical ordering (disposition, elaboration) and the linguistic formulation of a subject–specific matter or state of affairs. A genre is subject to linguistic norms relating to a particular national language.

The definition of genre offered by John Swales (1990:58) relates to genre as a category in LSP text linguistics and in literary studies. With reference to LSP (language for special purposes), he states:

> "A genre comprises a class of communicative events, the members of which share some set of communicative purposes. These purposes are recognized by the expert members of the parent discourse community, and thereby constitute the rationale for a genre. This rationale shapes the schematic structure of the discourse and influences and constrains choice of content and style."

2. Summarizing texts

All summarizing texts are *derived text genres* and depend on a previously existing *primary, original text*. Thus, summarizing texts are also relevant in the light of intertextuality. *Primary, original* text genres cover independent, self–contained texts which are an original contribution to a specific subject area or province of discourse. These include e.g. *monographs, research articles, doctoral theses, case studies, experimental reports, essays, patent applications, encyclopaedic articles for experts*, specialized *reference books*, etc. *Derived* text genres, on the other hand, are based on an underlying primary text and depend on its subject–matter, and its conceptual and terminological system. Thus, the author of a summarizing text is partly constrained in his/her scope of topic and linguistic means by the source text. Typical examples of *derived text* genres are *book reviews, abstracts* of *research articles* and *abstracts in abstracting journals*, the *summary* of a doctoral thesis ("Thesen"), the *conference report*, the *research/state of the art report*, the *review article*,

the *peer review*, the *testimonial* of a scholarly publication or manuscript, the *announcement* of a new publication in a catalogue issued by a publishing house, etc.

Derived texts based on *genres in oral (LSP) communication* include *minutes* of meetings and conferences, *student notes* of lectures and seminars, *personal notes* of conference papers and discussion contributions, *commentaries* on parliamentary debates, *press releases* on public debates and discussions in various institutions, etc.

An essential characteristic of summarizing text genres is that they select, evaluate, order and condense items of information according to their relevance for a particular subject or a particular purpose. Thus, the linguistic act (Sprachhandlung) of summarizing is a complex cognitive process which is composed of different cognitive operations. These are ultimately reflected in a text.

3. Summarizing as a cognitive and linguistic process

Summarizing paragraphs represent an important constituent of all derived texts. A text review will contain a passage on the special field in which the author is engaged, and a brief sketch of the content of the book under review. The minutes of a business meeting or a disputation of a doctoral thesis will report the main items and sum up the results of these communicative events. A peer review will concentrate on the originality of ideas or experimental findings communicated in a paper which has been submitted for publication.

The cognitive process of summarizing covers the following aspects:

- careful observation of details
- abstracting from irrelevant data
- evaluating data according to their relevance for the given topic or communicative purpose, also seen in the writer's or speaker's perspective
- generalization (distinction between generic and specific items)
- ordering items into a logical frame, with special reference to rhetorical techniques (time order, space order, cause and effect, comparison, contrast, etc. cf. L. Trimble 1985)
- condensing information
- lucidity of expression.

Egon Werlich (1988:42) offers a useful working definition of the text genre *summary*, without using the term genre. He prefers the general term text form. His definition reads:

> "*In a summary we present the information of a much longer text in much shorter reading or listening time* (emphasis – R.G.). Through a summary we can inform others about the

> contents of, for instance, a book or a long chapter in the length of only one short paragraph or even of only one sentence. The original text is translated into a new text.
> A summary can be based on a topic outline of a longer text. It can also start from concepts or text interpretation by which the topics of a long text can be easily ordered. (...) Its **analytical order** shows that the summary is an expository textform. It aims at **comprehension.**"

The criterion of comprehension and intelligibility is of utmost importance for each summarizing text. Transparency of thought and clarity of language are imperative.

The problem of condensing information in scientific texts was discussed by Lothar Hoffmann in 1988 (editor of the collection of papers entitled "Informationsverdichtung und Standardisierung in wissenschaftlichen Veröffentlichungen"). L. Hoffmann makes a number of interesting suggestions as to how information can be concentrated and redundancy be avoided in LSP publications. His source materials are encyclopaedic articles. Thus, careful attention should be paid to the macrostructure and the structure of paragraphs. Each paragraph should begin with a topic sentence, subheadings should be used when a new sub–topic is introduced and elaborated. Moreover, the title of a paper should correspond exactly to the macroproposition of the whole text. Finally, syntactic and semantic cohesives such as adverbials, modals and conjunctions should be used sparcely.

4. Teaching summarizing genres in courses of academic writing in present day English

The logical operation of summarizing texts in a foreign language presupposes experience in this cognitive performance in the mother tongue. A considerable number of students attending courses in English academic writing (at the Foreign Language Centre of Leipzig University) had little previous knowledge or none at all in this respect. Thus, the foreign language teacher must make the student aware of how to select and order items in summarizing texts, starting with the seemingly trivial genre of the *student's curriculum vitae*. The C.V. is an expository text form – in Werlich's terms – when composed as a coherent stretch of text and not as a table of chronological events in the student's life.

Undergraduate and even post–graduate students, however, were faced with difficulties when they were asked to write a German report on project work done in the classroom or a *conference report* intended for a specialist journal. Their main problem seemed to be

- how to select the relevant facts and to link them logically
- how to evaluate items of information

- how to establish the chronological order when reporting on parallel events as in the case of multiple sessions at an international conference
- how to use the appropriate speech acts and language functions (narration, description, argumentation, conclusion).

With regard to composing summarizing texts in English, the students will pass through a process of "awareness raising". They will be made familiar with the linguistic properties of well–formed summarizing texts.

The range of summarizing genres dealt with in the course of Academic Writing, among other texts, includes the following genres:

A) Abstracts

This genre is in fact the umbrella term of three subgenres

a) a *pre–text* ("unfinished", "promissory text")
   a text to be elaborated into a full text, mostly a conference paper
b) a *derived text*
   which is closely connected to the research article which has been written by the same author
c) a *derived text*
   which is physically separated from the research article to which it refers, and which may have been written by a different author. As a rule, it is published in an abstracting journal.

The summarizing genre abstract has been described in linguistic in–depth analyses, in guidelines for technical writers, and by international standards (cf. R. Gläser 1990, H.J. Meyer/H. Heidrich 1990, and Th.N. Huckin/L.A. Olsen 1992). All authors agree in that abstracts as *derived texts* must exactly correspond to the sequence of text segments in the macrostructure of the underlying research article. Thus, the first sentence will provide some background information on the state of the art, by way of introduction. The second sentence will relate to the specific topic, the problem under discussion, and the methodology applied. The third sentence may report on the major findings, experimental results or the confirmation of a hypothesis. The fourth sentence reflects the text segment 'conclusions' and perspectives for further research. Authors of handbooks on academic writing emphasize that the different functions of the individual sentences which constitute an abstract are clearly marked by the use of different tenses of the finite verb (present, present perfect in statements; simple past in narrative description).

B) Book reviews intended for specialist journals

A distinction is made between the academic book review and the book review for a popular journal, because the latter has particular stylistic features (a colourful heading, often an 'introductory hook' by way of an episode or an eye–catching item). The student is made familiar with the macrostructure of a book review and its content invariables, which constitute the whole *derived text*. These include: bibliographical data, the state of the art as background to the book under review; a brief note on the author (his/her previous work); *summary* of the books content; comment to, and evaluation of, theory, method, findings, and the innovative impact of the book under review. Optional content invariables are a note on the author's style and the quality of illustrations.

C) Conference reports

A report on a national or international workshop, symposium, conference or congress will inform the scientific community of the event under discussion and place it in the context of periodical conferences in the same field. The conference report includes the general theme of the conference, the name and institution of the organizers and, as a rule, mentions the number of participants. The student who is faced with writing a conference report may have difficulties in summarizing diversified topics of the conference papers read in parallel sessions, and of presentations in workshops and in poster–sessions. In this case, the writer/ reporter will decide on the basis of his/her subject competence to what extent new trends in scientific development could be observed at the given congress.

As a rule, summarizing passages in a conference report will relate to the plenary papers and keynote speeches. The report may conclude with an evaluative remark on the atmosphere of the conference, and its contribution to international research. Some conference reports end with a note on the next conference when following in a certain interval.

To sum up, the choice of genres and sub–genres which primarily have a summarizing function and thus occupy a prominent position in an English Academic Writing Course, is open–ended and will vary from term to term according to the students' needs.

5. References

Gläser, Rosemarie (1990):
"Das Abstract". In: R. Gläser: Fachtextsorten im Englischen. Gunter Narr Verlag, Tübingen, S. 117–130

Gläser, R. (1992):
"Die Arbeit mit Textsorten im fachbezogenen Fremdsprachenunterricht". In: Claus

Gnutzmann/Frank G. Königs/Waldemar Pfeiffer (Hrsg.): Fremdsprachenunterricht im internationalen Vergleich. Perspektive 2000. Verlag Moritz Diesterweg, Frankfurt am Main, S. 168– 189

Gläser, R. (1993):

"Textsortenvergleich im universitären Fremdsprachenunterricht". In: Klaus Morgenroth (Hrsg.): Methoden der Fachsprachendidaktik und –analyse. Deutsche Wirtschafts– und Wissenschaftssprache. Peter Lang, Frankfurt am Main – Berlin – Bern – New York – Paris – Wien, S.17–38

Hoffmann, Lothar (1988) (Hrsg.):

Informationsverdichtung und Standardisierung in wissenschaftlichen Veröffentlichungen. – Wissenschaftliche Zeitschrift der Karl–Marx–Universität Leipzig. Gesellschaftwissenschaftliche Reihe. Jahrgang 37, Heft 6, S. 552–565

Huckin, Thomas N. and Leslie A. Olsen (1991):

Technical Writing and Professional Communication. For Nonnative Speakers of English. New York etc., McGraw Hill, Inc. Second edition

Meyer, Hans Joachim/Hans Heidrich (1990):

"Unterordnung und Komprimierung von Aussagen." In: H.J. Meyer/H. Heidrich: English for Scientists. A Practical Writing Course. Verlag Enzyklopädie Leipzig, S. 18–26

Swales, John M. (1990):

Genre Analysis. English in academic and research settings. Cambridge University Press. Cambridge – New York – Port Chester – Melbourne – Sydney

Trimble, Louis (1985):

English for Science and Technology. A discourse approach. Cambridge University Press. Cambridge – London – New York – New Rochelle – Melbourne – Sydney

Werlich, Egon (1988):

"Summaries / Summarizing minutes". In: E. Werlich: Student's Guide to Text Production. Cornelsen Verlag, Berlin, S. 42–52

Human Summarizing – practice and models II

Rapporteur: Ines Busch–Lauer

Felix Friedrich:

Training of Reductive Text Learning Strategies

Felix Friedrich (Tübingen) reported on a prototypical training research of summarizing which relies on macrorules and structure–strategy training in expository text performed in an experiment within a population of 48 law students (1/2 term). Based on the core skills:

1. Identifying relevant information in a text,
2. Condensing the content of a text and
3. Handling text organization,

REDUTEX – a self–instructional program – was developed and tested both in a monothematic (law text) and multithematic context (law, biology, history).

In a posttest, subjects were asked to write a 300 word summary of the "near–to–subject matter" and then of the "distant–to–subject matter" text. A further 100–word summary writing was performed after an intervening period of time. The difference in the mean properties of reproduced main ideas was not significant (p=.14). "Practice–only" and "macrorule–and–structure" group performed equally well on "summarizing–without–access–to–the–target–text" whereas the combined training scheme showed slightly better trends while "summarizing–with–access–to–the–target–text".

Referring to the "near–to–subject matter" all patterns indicate superiority of "practice–only" strategies but with the "distant–to–subject matter" text general heuristics was dominating.

In the discussion, the plausibility of the presented explanation was questioned by the audience.

Usefulness of such training programs fostering follow–up research and necessity of feedback to students were underlined. Though summarization is not of proper interest to German law students this activity will be required for court case abstraction.

Edward Cremmins:

Valuable and Meaningful Text Summarization in Thoughts, Words, and Deeds

Edward Cremmins (Rockville, MD) reported on text processing practice and summarizing texts based on his long–standing experience as an abstractor, editor, translator and teacher. Using two research summaries of different nature (a scientific and a literature source) he elaborated on the natural thought process, the language skills and procedures that were used while preparing these documents.

Writing summaries is difficult primarily because of difficulties in text contents and in its organization the degree of comprehension, availability of text, audience, intended purpose, type of summary required, genre and text length. During the long process of learning to write summaries, initial instruction should be based on using narrative structures with a time–ordered format.

Structured materials covering experimental research in the natural sciences are far easier to be abstracted than the ideology–based texts in the humanities. No single item dominates in the abstracting process. Abstracting is a complex process being basically determined by domain–specific constraints, communicative purpose and experience of abstractor.

Training of Reductive Text Learning Strategies

Helmut Felix Friedrich

Deutsches Institut für Fernstudien an der Universität Tübingen

1 Research questions

This contribution describes a training approach which combines macrorules– with structure strategy–training to foster summarising expository texts. An empirical study examined the following questions:

1. How effective is this combined training approach?
2. Has a multithematic version of the training (practising with "near–to–subject matter" *and* "distant–to–subject matter" texts) a positive impact on post–training strategy transfer compared to a monothematic version?
3. How does the training (as a content independent heuristic) interact with domain specific prior knowledge?

2 Theoretical background

The skill to summarise expository texts can be an end of instruction in itself. In this study it is conceived as means to an end. As a learning strategy which promotes a better qualitative understanding and recall of the main ideas in the respective content domain. All in all, the task of summarising is a mean to stimulate those constructive mental activities which are highly relevant for text comprehension. Summarising tasks trigger the process of "imposing of meaning" (Resnick, 1987), the process of "text–reader–interaction" (Ballstaedt et al., 1981; Eigler et al., 1990).

2.1 Core components of efficient text summarisation

Nearly all theoretical and empirical studies (cf. Schnotz, Ballstaedt & Mandl, 1981; van Dijk & Kintsch, 1983; Winograd, 1984; Hidi & Anderson, 1986; Endres–Niggemeyer, Waumans & Yamashita, 1991) underline the importance of the following core skills for efficient summarisation:

1. Identifying relevant information in a text,
2. condensing the content of the text, and
3. handling text organization.

2.2 Training of the core components

These components are the object of several training studies. Some of them limit themselves to foster the skill to identify relevant information (e.g. Stevens, 1988). Others prefer to train semantic macrorules e.g. Hare & Borchardt, 1984). A third group of studies deals mainly with the training of structure strategies (e.g. Meyer, Young & Bartlett, 1989). All in all, these trainings seem to have positive effects on

(a) the acquisition of strategy knowledge,
(b) the acquisition of strategy mastery, and
(c) the comprehension and recall of content knowledge, especially main ideas.

Only limited evidence exists for far transfer effects of the training. A closer analysis of 12 training studies yields the following desiderata (Friedrich, in press):

1. In none of the 12 studies the *macrorule approach was combined with the structure strategy approach*. This is astonishing insofar as relevant theoretical and empirical work in this field emphasizes the interaction of semantic macrooperators and structure strategies (e.g. Schnotz, Ballstaedt & Mandl, 1981; van Dijk & Kintsch, 1983; Endres–Niggemeyer, Waumans & Yamashita, 1991).
2. There is only a limited amount of systematic experimentation in the *instructional design aspects* of the training, especially with respect to training conditions, like a multithematic training context, which foster the post–training transfer of the learned strategies.
3. The *interaction between domain specific prior knowledge and general strategies* is neglected. There is some evidence that training in general strategies can interfere with a prior knowledge driven mode of information processing (Lohman, 1986; Clark, 1990).

3 A combined macrorule– and structure strategy–training

As a prerequiste to answer the research questions a prototypical summarisation training – REDUTEX – was developed which combines the macrorules and structure strategy approach (Friedrich, in press). REDUTEX is a written self–instructional program to foster the application of macrorules (deletion, generalization, selection, construction) and the analysis of texts with the help of schemata (like GENERAL PRINCIPLE & EXAMPLE, CLASSIFICATION, QUESTION/ANSWER etc.). Finally, the two strategy components are combined in the following 5–step–heuristic:

1. Getting an overwiew over content and formal structure of the text.
2. Identification of the formal structure of the passages of the text.

3. Identification of the passages containing the information which presumably should be included in the summary.
4. Passagewise summarisation: Application of macrorules on selected passages.
5. Integration of the passage summaries to a text summary. This heuristic does not claim to be an empirically validated model of the interaction of macrorules and structure strategies. It has more the status of an educationally motivated simplification of a complex process.

4 Evaluation of the training

4.1 Methods section

Subjects:

N = 48 law students (1st/2nd term) randomly assigned to the various treatments.

Design:

3x2x2 ANOVA with repeated measurement on the last factor.

Independent variables:

1. *Version of training:*

   (a) "practice only" control group (PRACTICE ONLY),
   (b) macrorule only–training (MACRO),
   (c) combined macrorule– and structure strategy–training (MACRO&STRUCTURE).

2. *Context of training:*

   (a) monothematic context (MONO): practising only with law texts,
   (b) multithematic context (MULTI): practising with law, biology and history texts.

3. *Context of application (within subjects factor):* In the posttest the trained strategies were

   (a) first to apply to a "near–to–subject matter" text (NEAR) and then
   (b) to a "distant–to–subject matter" text (DISTANT).

Dependent variable:

Ratio between the number of main ideas reproduced in a summary and the total number of main ideas in the respective target text.

Task conditions:

Summarising the texts in the posttest first with and then without access to the target text.

Procedure:

Prior knowledge of the subjects with respect to the target texts (to be summarized in the posttest) was assessed by some open questions in a pretest session. The training with the experimental versions of REDUTEX took place a week after the pretest. In the posttest (one week after training) subjects had first to write (with access to the target text) a 300–word–summary of the "near–to–subject matter" (about "Theories of Justice") and then of the "distant–to–subject matter" text (about "Problem Solving"). Both texts were of equal length (ca. 900 words) and had the same top level structure (DESCRIPITION). After some intervening tasks a 100–word–summary was to write of both texts without access to the target texts (recall condition). (For further details, scoring etc., see Friedrich, in press).

4.2 Results and discussion

4.2.1 How effective is the combined training approach?

The following pattern of results would favor the combined training approach:

MACRO&STRUCTURE > MACRO > PRACTICE ONLY.

Under the "summarising with access to the target text"–condition the following pattern was observed:

PRACTICE ONLY(.75) > MACRO&STRUCTURE(.72) > MACRO(.66).

The differences in the mean proportions of main ideas reproduced in the summaries are not significant (main effect "version of training": $p = .14$).
Under the "summarising without access to the target text"–condition the following pattern emerged:

PRACTICE ONLY (.44) = MACRO&STRUCTURE (.44) > MACRO (.36).

In this case the main effect "version of training" is significant ($p = .03$). Since there exists a significant interaction (cf. 4.2.3), this main effect should be interpreted with caution. Both patterns favor the structure–strategy training approach over the macrorule training approach which showed under all conditions the smallest proportion of main ideas reproduced in the summaries. But even the structure strategy approach is less effective than the practice only condition.

4.2.2 Does the multithematic training context promote strategy transfer?

If the multithematic training context promotes strategy transfer to a "distant to–subject matter"

domain, a significant interaction "context of training x context of application" is expected. No patterns of results consistent with this expectation emerged, neither under the "summarisation with access"– nor under the "summarisation without access to the target text"–condition. An explanation may be that the training of an unfamiliar strategy in an unfamiliar content domain may cause cognitive overload and deteriorate learning.

4.2.3 Does the training interact with prior knowledge?

If there exists an antagonistic relationship between a knowledge driven and a heuristic driven mode of summarising, the ANOVA should yield a significant interaction "version of training x context of application": The combined training approach (MACRO & STRUCTURE) – a content independent heuristic – should be superior in summarizing the "distant–to–subject matter" text, whereas in the case of the "near–to–subject matter" text PRACTICE ONLY should be superior.

This interaction shows a trend in the expected direction, but did not reach significance ($p = .14$) under the "summarising with access to the target text"–condition. For the "near–to–subject matter" (A) and the "distant–to–subject matter" text (B) the following patterns of results were observed:

(A) PRACTICE ONLY (.72) > MACRO&STRUCTURE (.67) > MACRO (.59)
(B) MACRO (.62) > PRACTICE ONLY (.61) = MACRO&STRUCTURE (.61).

Under the condition "summarising without access to the target text" this interaction reached significance ($p = .03$). The patterns of results were as follows:

(A) PRACTICE ONLY (.48) > MACRO&STRUCTURE (.40) > MACRO (.36)
(B) MACRO&STRUCTURE (.43) > PRACTICE ONLY (.35) > MACRO (32).

All patterns indicate that in the case of a "near–to–subject matter" text the individual strategies (PRACTICE ONLY) are superior to the content independent heuristic (MACRO & STRUCTURE), whereas in the case, a "distant–to–subject matter" text is to summarise under recall conditions, the general heuristic is superior. This interaction indicates an antagonistic relationship between a prior knowledge driven and a heuristic driven mode of text summarising. This interpretation is corroborated by a small but significant prior knowledge difference in the pretest: The law students possess more prior knowledge with respect to the "near–to–subject matter" text about theories of justice than with respect to the "distant–to–subject matter" text about problem solving (mean of 1.8 vs. mean of 1.1 points; $t = 2.21$, $df = 47$, $p=.02$, one tailed, paired t–test).

The results of this study point in two directions:

1. With respect to the general discussion about strategy training the results underpin the necessity to consider the "costs" of strategy training (Clark, 1990) which can result from a conflict between a prior knowledge driven mode and a heuristic driven mode of information processing.
2. With respect to the theme of this conference, the results show that heuristics for text summarizing – especially those heuristics which rely on structure strategies – can be helpful, if someone has to summarize texts in an unfamiliar content domain.

References

Ballstaedt, S.–P., Mandl, H., Schnotz, W. & Tergan, S.–O. (1981).
Texte verstehen – Texte gestalten. München, Wien, Baltimore: Urban & Schwarzenberg.

Clark, R.E. (1990).
When teaching kills learning: Research on mathemathantics. In H. Mandl, E. De Corte, N.S. Bennett & H.F. Friedrich (Eds.), *Learning and instruction: European research in an international context* (Vol. 2.2, pp. 1–22). Oxford, UK: Pergamon.

Eigler, G., Jechle, Th., Merziger, G. & Winter, A. (1990).
Knowledge and text production. In H. Mandl, E. De Corte, N.S. Bennett & H.F. Friedrich (Eds.), *Learning and instruction: European research in an international context* (Vol. 2.2, pp. 341–356). Oxford, UK: Pergamon Press.

Endres–Niggemeyer, B., Waumans, W. & Yamashita, H. (1991).
Modelling summary writing by introspection: A small–scale demonstrative study. *Text, 11*, 523–552.

Friedrich, H.F. (in press).
Training und Transfer reduktiv–organisierender Strategien für das Lernen mit Texten. Tübingen: Deutsches Institut für Fernstudien.

Hare, V.Ch. & Borchardt, K.M. (1984).
Direct instruction of summarization skills. *Reading Research Quarterly, 20*, 62–78.

Hidi, S. & Anderson, V. (1986).
Producing written summaries: Task demands, cognitive operations, and implications for instruction. *Review of Educational Research, 56*, 473–493.

Lohman, D.F. (1986).
Predicting mathemathanic effects in the teaching of higher–order thinking skills. *Educational Psychologist, 21*, 191–208.

Meyer, B.J.F., Young, C.J. & Bartlett, B. (1989).
Memory improved. Reading and memory enhancement across the life span through strategic text structures. Hillsdale, NJ: Erlbaum.

Resnick, L.B. (1987).

Instruction and the cultivation of thinking. In E. De Corte, H. Lodewijks, R. Parmentier & P. Span (Eds.), *Learning and instruction: European research in an international context* (Vol. 1, pp. 415–442). Oxford: Pergamon Press.

Schnotz, W., Ballstaedt, S.-P. & Mandl, H. (1981).

Kognitive Prozesse beim Zusammenfassen von Lehrtexten. In H. Mandl (Hrsg.), *Zur Psychologie der Textverarbeitung: Ansätze, Befunde, Probleme* (S. 108–167). München, Wien, Baltimore: Urban & Schwarzenberg.

Stevens, R.J. (1988).

Effects of strategy training on the identification of the main idea of expository passages. *Journal of Educational Psychology, 80*, 21–26.

van Dijk, T.A. & Kintsch, W. (1983).

Strategies of discourse comprehension. New York: Academic Press.

Winograd, P.N. (1984).

Strategic difficulties in summarizing texts. *Reading Research Quarterly, 19*, 404–425.

Valuable and Meaningful Text Summarization in Thoughts, Words, and Deeds

Edward Cremmins

Independent Researcher
10075–6 Windstream Drive
Columbia, Maryland, USA 21044

The diverse types and characteristics of human–produced summaries of text are described, and the natural thought processes, language skills, and procedures that were used while preparing several examples of "research summaries" for intelligent communication of text are discussed.

Recommendations for sounder natural thinking while writing clearer summaries include:

- Select a logical format for composing the summary and adhere to the required order for representing the relevant information into it.
- Look for semantic cues as to the primary aboutness of the text.
- Suspend final judgement on what information is relevant for inclusion in the summary until the full text has been scanned.
- Do not misrepresent elements of text through careless extraction from it, nor overemphasize the significance of certain relevant information at the expense of other equally or more relevant information through inadequate reading of the text of the basic document or inattentive listening to spoken text.
- Attempt to substitute more precise words for any unclear or ambiguous ones that are extracted from the material being summarized. The substitutions should be made only when the summarizer is convinced that there is not the slightest possibility that the author's meaning will be distorted in the process.
- Appropriate use should be made of such tools for condensing or gisting as outlining, notetaking, indexing, deleting, inferring, selecting or creating topic sentences, and substituting generalized terms for lists of things or actions.

Through these and other ways of exercising the natural thought processes while summarizing not only will well–structured and meaningful summaries be written, but also the summarizers themselves will improve their ability to comprehend and gist information and increase their practical knowledge in such subject areas as semantics, syntax, grammar, usage, reading, listening, writing, editing, and revising.

Human Summarizing – practice and models III

Rapporteur: Helmut Felix Friedrich

The contribution of W. *Schnotz* and Th. *Zink* ("Macrostructure and mental models in summarizing learned information") had to be cancelled because both authors couldn't join the workshop.

The discussion about the contribution of Harold *Borko* and Brigitte *Endres–Niggemeyer*:

A naturalistic model of abstracting

focused on the following points:

1. Comparison between the blackboard model presented by the authors and other systems

   One of the main differences between the Endres–Niggemeyer model and other systems, for example SUSY, seems to be, that the former has a "dynamic understanding"–component (multi–parse feature). "Dynamic understanding" means in this context, that the level of processing is regulated in a flexible manner between a more shallow and a more deep mode of processing, depending of comprehension problems coming up during the reading process. SUSY can be characterized as a one–parse system, which treats the comprehension process in a fixed linear sequence without regulation of "levels of processing".

2. Consistency of expert behavior

   The question was raised, if there exists some agreement between the subjects (6 experts in abstracting) not only on the strategic level, but also with respect to some others features (time needed to solve the task, quality, length, structure and other surface features of the abstract). The authors saw the similarity between the different experts in the intellectual tools and strategies, the experts used to solve the task: nearly each of the (about 150) tools could be identified by help of the thinking aloud procedure in the working process of 5 of the 6 subjects. But with respect to other aspects of process (time and effort invested) and product (features of the abstracts) there were great differences between the experts.

3. The meaning of "strategy"

   It was criticized that the multitude of strategies in the model is of rather heterogenous

character. Some of them, as for example the tool "question", are rather complex activities with a strong covert cognitive component, others, as for example "underlining", are overt behavior components, which may be in the best case indicators of some underlying cognitive acticity/strategy, but not the strategy itself.

4. "Implementability" of the model

In responding to some questions concerning this point, B. Endres–Niggemeyer stated, that some of the components of her model are implemented in other systems, but that other components, for example the "All Purpose Knowledge Processing"–component, are still far from implementation.

5. Dependency of the model from a specific working environment

The question was raised, if the model is restricted in its generalizability to a certain kind of working environment, in this case a "low tech" (a paper & pencil) environment, and if a high tech environment (a workbench with cut & paste facilities) would not necessitate adaptions in the model. The discussion about this point was controversial. Some supposed a marked influence of the working environment on cognitive processing, while others thought this influence to be restricted to peripheral aspects, for example the speed of re–arranging a draft, whereas the main "bottle neck" in abstracting – human cognitive processing – is not changed by the working environment.

A Naturalistic Model of Abstracting

Brigitte Endres–Niggemeyer, Hannover

(with Harold Borko)

1. Introduction

Abstracting, or abstract writing, is a professional specialization of text summarization for bibliographic informations systems, dealing with long documents which are scientific or technical most of the time. The process model of abstracting presented here is expected to explain how abstractors go about (HUTC87). It is a naive qualitative model of expert performance (NORM83, WELD90).

2. Modelling Methods

Modelling the abstracting process means to develop a grounded theory and a naturalistic model (GLAS67, DIES71, LINC85) or, following the KADS terminology (WIEL91), a conceptual model of abstracting. 36 abstracting processes of 6 experts from Germany and the United States have been recorded on tape via thinking–aloud protocols, transcribed and interpreted (HAYE80, ERIC84). Interpretation was mainly driven by the model of text understanding and summarizing of KINT83 and, to a lesser extent, by the process model of text production by HAYE80. The experts have expanded the summary writing knowledge of these models with their personal, mostly experiental know–how.

3. The Naturalistic Model of Abstracting

As a result of empirical modelling, we know better how expert abstractors organize their working processes, which intellectual tools (standard strategies) they use, and how successful natural working contexts (e.g., steps or moves comprising several steps) look like (ENDR92).

Planning, Control, and General Literacy
plan: State what you are going to do next.
question: Formulate questions in order to answer them from input, e.g., from your document.
inference: Infer implied knowledge.
underline: Underline a text passage.

Information Acquisition

start–explore: Begin to explore the document.
explore: Get a document meaning item.
hold: Keep an information unit.
first: Look at the beginning of text organization units.
browse: Explore the document by normal reading and understanding.
read: Read sequentially.

Relevance Assessment

relevant–say: State a topic item in your own words.
relevant–unit: An information item is recommended as relevant by its position (beginning or end).
relevant–topic–sentence: Topic sentences are relevant.
relevant–call: Meaning items that are called by the text theme are relevant.
*relevant–texthint:*Exploit textual hints to decide about the relevance of information items.

Fig. l: A choice of abstracting tools: Strategies used in fig.2

3.1 Organizational Principles of Abstracting Processes

1. Expert abstractors work step by step. In every step, several strategies cooperate. A leading strategy defines the overall goal of the working step (e.g., to acquire information, or to produce a text unit) and links the step to the working plan. All strategies fit into an ordered intellectual toolbox.
2. The working processes have a general pattern of knowledge processing: Knowledge items from the original text are isolated and either dropped from processing or moved into the target representation, passing through intermediate products. Beyond this basic pattern of knowledge processing, experts follow a loosely coupled experiental working plan. Typical planning units are the exploration of a table of content, or the production of a statement.
3. Professional expertise is embedded into general skills. The experts invest metacognitive activity (general monitoring and self–steering – FLAV81, GARN87), they draw heavily on general literacy skills (reading, writing, and thinking), and they dispose of control strategies for working processes.
4. Task–oriented memory areas (schemata) structure the text knowledge that is used during a working process. By their built–in views the schemata impose on their data the right structure for the intended type of processing, and exclude features that don't fit. Three source text oriented schemata are needed: surface text, document scheme, and document theme. Since output is smaller in in size, one can do with one product memory that contains different subareas.

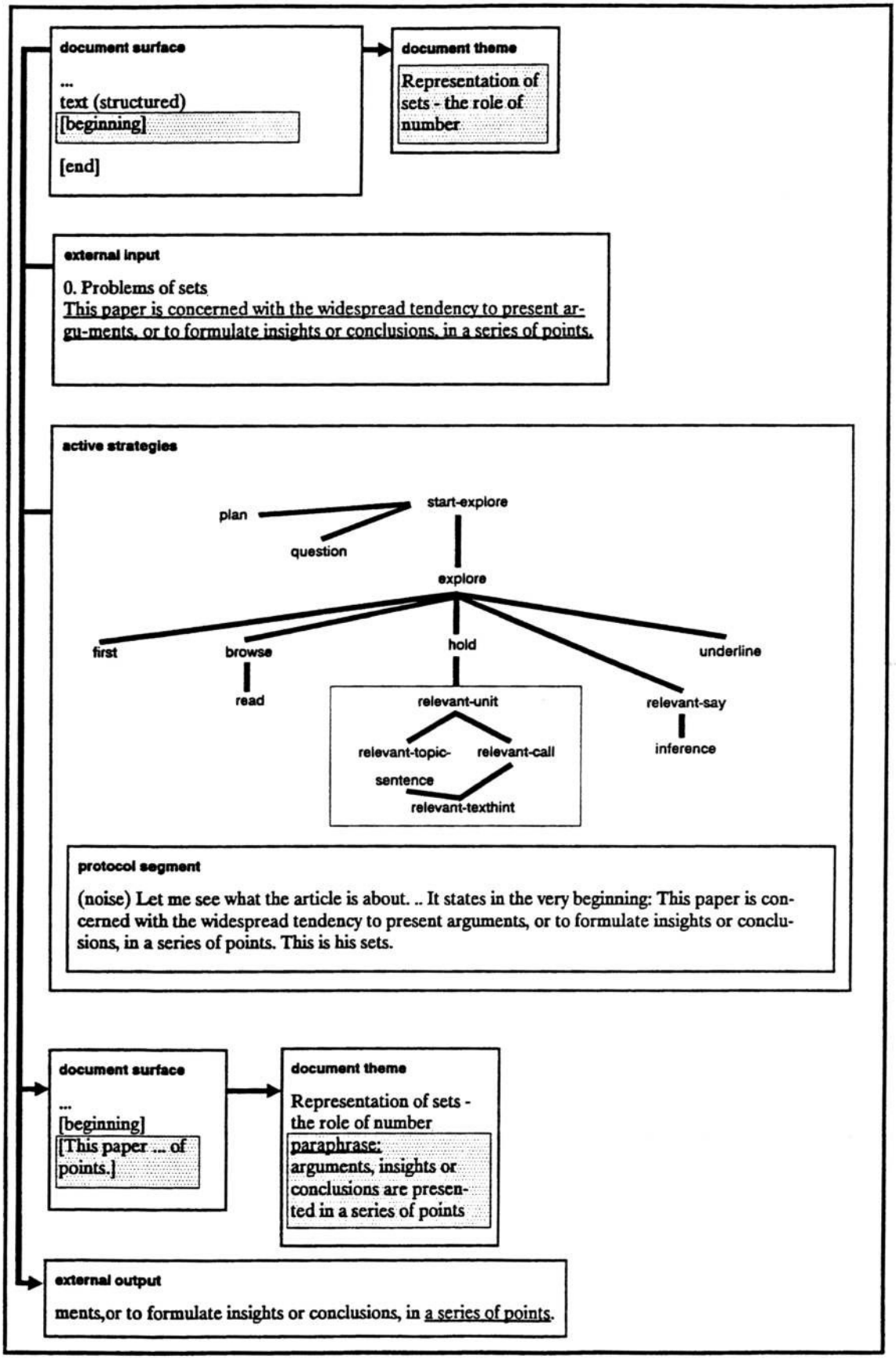

Fig. 2: A Sample Abstracting Step: "Let me see what the article is about"

3.2 The Intellectual Toolbox

With its 453 abstracting strategies (see sample in fig.1), the intellectual toolbox represents an abstractor's methodological expertise. As elsewhere, expertise is composed by different types of skills (STEEL90).

The core abstracting expertise consists of two main groups of tools: strategies for information acquisition, and presentation strategies. The former includes tools for interaction with the document (e.g., navigation, and elaboration), for dynamic task–oriented information seeking, and for holding relevant items. The latter contains tools for informational upgrading of material, for abstract construction, and for formulation.

3.3 An Individual Working Step

Whereas the intellectual toolbox shows how the expertise of abstractors is composed, individual abstracting steps help to understand how strategies cooperate successfully in natural contexts. Fig.2 displays a typical abstracting step. In the top windows, input data is represented, the process description figures in the middle, and output is presented at the bottom. The process description area in the middle presents the current segment of the thinking–aloud protocol, and a tree–like structure that characterizes how abstracting strategies cooperate.

Let's observe what happens. Our abstracting step in fig.2 is the third of the respective working process. From the two preceding steps the abstractor knows in particular the document title, i.e., a first formulation of the document theme (cf. fig.2, memory areas in top of the display; for strategy definitions, see fig.1). Now he is going to find out what the document is about.

In the current working step, the abstractor starts out to explore the document text (strategy *start–explore*), stating his plan and asking his standard "aboutness" question: "Let me see what the article is about ..." (strategies *plan* and *question*). He begins the first exploration step (strategy *explore*).

Reading behaviour is basic (strategies *browse* and *read*), from the thinking–aloud protocol we learn what has been read. The abstractor is aware of his promising position "in the very beginning" of the document (strategy *first*). He underlines "a series of points" (strategy *underline*, see external output area in fig.2) and states "This is his sets." (strategies *relevant–say* and *inference*). We conclude from both underlining and connecting to the title that the abstractor keeps for later use what he has read (strategy *hold*).

Four frequent relevance assessment strategies account for his reasons:

- *relevant–unit*, because the beginning of a unit is a well–known favourite place of topic sentences
- *relevant–texthint*, because the current statement is introduced by a typical indicator phrase of topic sentences ("This paper is concerned with ...")
- *relevant–call*, because the abstractor is able to attach the current statement to the document theme ("This is his sets.")
- *relevant–topic–sentence*, because topic sentences are notoriously relevant for abstractors.

In addition, the abstractor reformulates the statement under consideration with his own words (strategy *relevant–say*). This is a frequent form of text elaboration.

The newly acquired topic sentence shows up in the document theme area. It has been linked to the title with a paraphrase relation.

4. Empirical Design of a Simulation System

An empirical model is not only useful for its own sake. In the KADS methodology, the "conceptual model" prepares the "design model" that guides implementation. Consequently, an inductive system design starts out by a careful reconstruction of observational units, e.g., intellectual tools, or working steps, under the roof of an appropriate system model. Since the units of empirical observation are organized in their natural context, it must be possible to win an overall architecture for intact empirical processes.

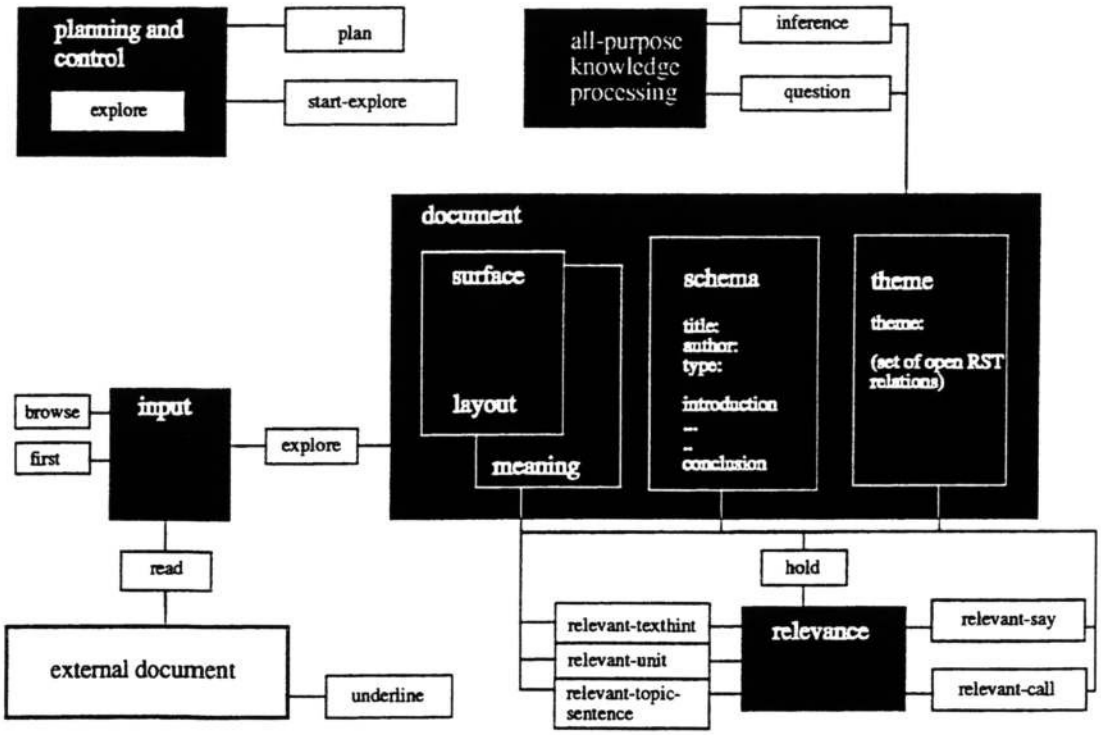

Fig.3: Blackboard View of the Sample Abstracting Step

The conceptual model supports the following features of an implemented simulation system:

- working steps as basic units of activity, with a planning and control structure behind
- rather independent small–scale agents in considerable quantities, namely the intellectual tools
- a blackboard–type communication and cooperation of agents
- a dynamic text representation that emphasizes document structure

Figure 3 presents our sample abstracting step in an implementation–oriented blackboard view (ENGE88, JAGA89). There, intellectual tools appear as agents that assemble around dedicated blackboards. E.g., information acquisition agents define and check the input of interest on the input board; agents of planning, control, and metacognition dip into the working step through the planning board. As usual, agents may activate each other by messages. While the document board and the planning board keep information beyond a working step, other boards are strictly local. The relevance board, for instance, can be wiped after every single relevance decision.

On the document board, the three tasks–specific general views of the document (see above section 3.1 and fig.2) are assembled during the working process. The scheme view and the thematic view use RST–like relations (MANN88) to structure meaning units. Individual agents may develop own task–oriented text views from document board representations. For instance, the strategy *relevant–texthint* analyses an interesting text passage into an indicator phrase and its scope.

Since other working steps look quite similar and at least some intellectual tools look implementable, the system sketch won from the conceptual model allows an explorative implementation. In spite of shared theoretical background (KINT83) and some shared features (e.g., the relevance strategies) the "grounded" system design proposed here differs strikingly from its nearest neighbour, the SUSY system (FUM82/85). Most of the difference is due to empirical observation.

5. References

(COLI82) COLING-82:
Proc. of the 9th Int. Conference on Computational Linguistics. Prague.

(DICK81) Dickson, W.P. ed. (1981):
Children's Oral Communication Skills. New York: Academic Press.

(DIES71) Diesing, Paul (1971):
Patterns of Discovery in the Social Sciences. Chicago: Aldine Atherton.

(ENDR92) Endres-Niggemeyer, B. (1992):
Abstrahieren, Indexieren und Klassieren. Ein empirisches Prozeßmodell der Dokumentrepräsentation. Hab. Konstanz, Fak. für Verwaltungswissenschaft.

(ENGE88) Engelmore, R; Morgan, T. eds. (1988):
Blackboard Systems. Wokingham: Addison-Wesley.

(ERIC84) Ericsson, K.A.; Simon, H.A. (1984):
Protocol Analysis: Verbal Reports as Data. Cambridge MA: MIT Press.

(FLAV81) Flavell, J.H. (1981):
Cognitive Monitoring. 35-60 in (DICK81).

(FUM82) Fum, D.; Guida, G.; Tasso, C. (1982):
Forward and Backward Reasoning in Automatic Abstracting. 83-88 in (COLI82).

(FUM85) Fum, D.; Guida, G.; Tasso, C. (1985):
Evaluating Importance: A Step Towards Text Summarization. 840-844 in (IJCA85).

(GARN87) Garner, R. (1987):
Metacognition and Reading Comprehension. Norwood NJ: Ablex.

(GENT83) Gentner, D.; Stevens, A.L. eds. (1983):
Mental Models. Hillsdale NJ: Erlbaum.

(GLAS67) Glaser, B.G.; Strauss, A.L. (1980):
The Discovery of Grounded Theory: Strategies for Qualitative Research. 11. ed. 1980. New York: Aldine Atherton.

(GREG80) Gregg, L.W.; Steinberg, E.R eds. (1980):
Cognitive Processes in Writing. Hillsdale NJ: Erlbaum.

(HAYE80) Hayes, J.R; Flower, L.S.(1980):
Identifying the Organization of Writing Processes. 17-35 in (GREG80).

(HUTC87) Hutchins, W.J. (1987):
Summarisation: Some Problems and Methods. 151-173 in (INFO87).

(IJCA85) IJCAI-85:
Proceedings of the 9th International Joint Conference on Artificial Intelligence. Los Altos CA: Kaufmann.

(INFO87) Jones, KP. ed. (1987):
Informatics 9: Proceedings of a Conference. London: Aslib.

(JAGA89) Jagannathan, V.; Dodhiawala, R; Baum, L.S. eds. (1989):
Blackboard Architectures and Applications. San Diego: Academic Press.

(KINT83) Kintsch, W.; Dijk, T.A. van (1983):
Strategies of Discourse Comprehension. Orlando FLA: Academic Press.

(LINC85) Lincoln, Y.S.; Guba, E.G. (1985):
Naturalistic Inquiry. Beverly Hills CA: Sage.

(MANN88) Mann, W.C.; Thompson, S.A. (1988):
Rhetorical Structure Theory: Toward a Functional Theory of Text Organization. Text 8(1988)3: 243–281.

(NORM83) Norman, D.A. (1983):
Some Observations on Mental Models. 7– 14 in (GENT83).

(STEEL90) Steels, L. (1990):
Components of Expertise. AI Magazine, Summer 1990:29–49.

(WELD90) Weld, D.S.; Kleer, J. de (1990):
Readings in Qualitative Reasoning about Physical Systems. San Mateo CA: Morgan Kaufmann.

(WIEL91) Wielinga, B.J.; Schreiber, A.T.; Breuker, J.A. (1991):
KADS: A Modelling Approach to Knowledge Engineering. Amsterdam: Univ. of Amsterdam, KADS–II/Tl. l/PP/UvA/008/1.0.

Automatic Summarizing – implementation and systems I

Rapporteur: Harold Borko

Udo Hahn:

Concept–Oriented Summarizing in the Text Condensation System – TOPIC

The architecture of the TOPIC text summarization system consists of a word parser that operates on a full–text database. Effective summarization requires *text cohesion* to reflect inter–sentence relationships and *text coherence* which deals with the thematic structure of the discourse. Three types of salient concepts are computed: Dominant Frames, Dominant Slots, and Dominant Fillers. These concepts are combined to make up a topic description which leads to the generation of text graphs. The graphs capture the salient features of the different summary applications such as abstracts, documents, and fact retrieval.

Six desiderata for future text condensation systems were discussed including procedures for combinig graphs, developing formal models of summarization as well as criteria for the evaluation of results. It was pointed out that TOPIC is a framework that can be used to experiment with different types of summarizations.

Ellen Riloff:

A Concept–Based Approach to Domain–Specific Text Summarization: A Proposal

Text summarization is similar to text classification in information retrieval in that both are based on information extraction in order to identify relevant portions of a text that should be included in a summary of that text. The system uses a "training corpus" of representative texts that has been classified by domain users as sets of relevant and non–relevant documents.

For each set a sentence analyzer algorithm called CIRCUS generates conceptnodes which contain slots into which specific types of information extracted from sentences are abstracted. For the document domain dealing with terrorism, the concept nodes and case slots consist of damage, weapons, perpetrators, victims, etc. Each case contains domain specific information extracted from a single sentence. Typically, only a subset of the cases would be deemed relevant, and only the relevant cases would be used to generate a summary of the text.

Comments: These two papers of Hahn and Riloff present different procedures of natural language analysis and text summarization. The TOPIC system described by Hahn relies on parser algorithms to create text graphs which can be used for different summary applications. Riloff's approach uses statistical techniques to automatically derive relevant concept nodes from a "training corpus" and then applies these indicators to new documents to extract information for use in text classification, information retrieval and text summarization. Both papers open up avenues for further research.

Concept–Oriented Summarizing in the Text Condensation System TOPIC

12 Claims and 6 Desiderata for Design

Udo Hahn

CLIF Computational Linguistics Research Group
Freiburg University
D–79085 Freiburg, Germany

This contribution to the Dagstuhl seminar on *"Summarising Text for Intelligent Communication"* serves two purposes. First, it summarizes our experience with the development of the text condensation system TOPIC in terms of 12 claims. All of them relate to methodological decisions that still seem valid even though a decade has passed since TOPIC's initial design was set up. Second, we consider various gaps in the TOPIC research framework and elaborate on 6 desiderata for future text condensation systems.

12 Claims about the Proper Design of a Text Condensation System

The TOPIC system has been documented in a series of publications [1–7]. Distinguishing between those design decisions which relate to natural language understanding and text summarization issues the following claims form the methodological substrate of our experience with the prototype (still alive and running!):

PARSING AND TEXT UNDERSTANDING

1. Clearly, the ability to produce text summaries evidences an agent's understanding of a source text, i.e. text summarization can be considered a litmus test for the methodological premises generally assumed to hold within the paradigm of *knowledge–based language processing*. Therefore, any text summarization system must be anchored in a powerful (e.g., KLONE–type) knowledge representation system. Text summarization is infeasible (or simply provides a poor level of output) when only linguistic processing (with some rudimentary form of semantics, if any) is performed.

2. The complexity of text understanding is kept manageable (in fact, it is reduced when

compared to alternative designs) when linguistic processing and conceptual inferencing operate in a carefully balanced manner. We might even go a step further and argue for a *strongly interactive* language processing model involving a *heterarchic* control structure as being particularly suited from a computational point of view, though one might dabate the issue whether interactivity should be organized in a strong or weak manner. However, neither a syntax–first nor semantics–first approach is appropriate when it is conceived as a fixed, built–in control structure.

3. Not only a close but also a *flexible* form of interaction between the linguistic and the conceptual processor has to be assured in terms of the dynamic, task–specific configuration of communication channels at run–time, i.e., in the course of text analysis. A suitable and effective architecture (as the one investigated in the TOPIC framework) consists of a *lexically distributed* (word–expert–style) parser cooperating with a *conceptually distributed* (frame–style) terminological knowledge base. An appropriate instantiation and – due to concept learning mechanisms – structural enhancement of that domain kowledge base constitutes a text knowledge base, the final outcome of proper parsing and text understanding processes.

4. Summarization efforts are only reasonable for sufficiently "long" texts (talking about summaries of 2–sentence texts is but a nice joke). But if longer texts (say, those running to at least 50 sentences) are processed, a suitable *text grammar/parser* has to be supplied [8–12]. Only such a device is capable of dealing with

   - *text cohesion* phenomena reflecting immediate inter–sentence relationships at the local level of textuality ((pro)nominal anaphora, ellipsis, etc.). If these phenomena aren't resolved, semantically *invalid* text knowledge bases will necessarily be generated;

   - *text coherence* phenomena applying to paragraph–level relationships at the global level of textuality (i.e., wellformedness constraints holding for valid argumentation lines, topic development patterns (thematic progressions, etc.), proper use of tenses, etc.). If those phenomena aren't accounted for, *understructured* text knowledge bases lacking pragmatic cues of high–level discourse organization will arise.

5. Non–toy–domain text understanders will never get the whole story. Underspecification of linguistic and domain knowledge as well as the occurrence of ungrammatical natural language utterances are inherent to any form of realistic natural language processing and thus require *robust* processing facilities. As a consequence, commit yourself and your

system to a tolerable degree of imperfection. Parsing real–world texts will continue to be *partial* parsing for a long while. Therefore, parsing devices must be tuned appropriately.

6. This skeptical appraisal of the current state–of–the–art of text understanding methodology needs to be supplemented by the definition of pragmatic research strategies in order to carefully advance the field. If full understanding in non–toy domains is illusionary in medium terms serious feasibility judgments with respect to less ambitious tasks should pave the way to presently solvable problem classes. As an attempt to synchronize evident methodological limitations with non–excessive requirements inherent to the envisaged application we consider the automatic generation of indicative abstracts, still a fairly advanced goal though, perhaps, at the lower end of what many people would expect from sophisticated NLU systems. As a matter of fact, one does not have to supply full–blown, fine–grained linguistic and conceptual representation structures if indicative abstracts are the desired target structure. Instead, one may concentrate on *nominal* structures as designators of the main themes at the linguistic level (leaving many critical aspects of full sentence parsing aside) and supply primarily *taxonomic–terminological* knowledge to reason about the "relevant" concepts at the domain knowledge level, thus neglecting major portions of assertional domain knowledge.

AUTOMATIC SUMMARIZATION, ABSTRACTING, AND TEXT CONDENSATION

1. Keep in mind the lesson that can be learned from the information science department next door: There are various forms of abstracts, each serving different information needs and purposes (e.g., indicative, structured, informative, critical abstracts, selective reviews). So any abstracting system must provide explicit mechanisms (processing strategies, parameters etc.) to *scale its (future) performance*. Even if the particular type of abstract to be generated is already agreed upon there exists no canonical, a priori valid format, size or knowledge granularity that will turn out to be appropriate for many different information retrieval situations, users, etc. The paradigm of *text condensation* is particularly sensitive to this observation. It requires that texts be summarized with *variable degrees of information granularity* (this idea had originally been coined "cascaded abstracts" [13]).

2. Text summarization is a transformation process operating at the knowledge level of text propositions, i.e., text knowledge bases. Major methodological criteria for concept–oriented summarization relate to

    - *connectivity* patterns of the underlying concept graph;

    - individual *weights* attached to concept descriptions (indicative of conceptual

salience (relevance) in terms of the frequency of referring to (not literal mentioning of!) the corresponding lexical items in the text being analyzed);

- *abstraction hierarchies* in frame representations with regard to
  - generalization/specialization relations holding between slot fillers and slots as well as associative relations (many of which are PART–OF) between slots and their associated frame;
  - hierarchical taxonomic relations (ISA and INST–OF) among the concepts involved.

3. The representation of a summary in terms of a text condensate is neither a linguistic description nor a simple conceptual conglomerate but a respresentation structure on its own we here refer to as *text* or *condensate graph*. Text graphs combine standard representation features inherited from their underlying knowledge representation model (well suited to represent the summary's content) with additional link features providing reference to plain text (thus constituting the basis for text passage retrieval) and to factual knowledge (data missing in the summary that can be retrieved from remote text knowledge bases). The vertices of text graphs are therefore constituted by single concepts or composite concept graphs, even (portions of) knowledge bases, and multi–media objects (texts, pictures, graphics, etc.). Their edges consist of
   - taxonomic relations (ISA, INST–OF, PART–OF, etc.),
   - hypermedia relations linking text passages with text knowledge portions and related graphical items (tables, pictures, etc.).

   Experiments with extended versions of text graphs (including co–authoring and group argumentation applications [15–18]) show their potential to serve as an even more general medium for text knowledge structuring. In particular, we have augmented the basic set of summary–oriented text graph relations outlined above by
   - assertional, RST–style, causal or temporal relations for deep text knowledge extraction,
   - argumentative, Toulmin–style relations (evidence, warrant, backing, qualification), e.g., for text production,
   - annotative relations for knowledge base synthesis (version–of, in–conflict–with,

subsumption/redundancy, etc.).

4. Text graphs form the backbone for the *visualization* of complex text knowledge bases and graphical, hypertextual interaction, i.e., *navigation* through text graph collections [14]. We thus advocate a direct navigational style to explore condensate information (with all those "getting–lost" problems one usually encounters in any complex hypertext environment!). Natural language summaries play a rather marginal (actually, no) role in the retrieval mode, but may perfectly fit in a kind of report generation mode after successful exploration.

5. Access to the system's text graphs is inherently coupled to facilities for *dynamically tailoring* abstracts. This allows *user–centered adaptation* of the granularity of the summary according to the user's current state of knowledge, interests, information requirements, etc.

6. Text summaries (abstracts) constitute a single, *specialized* feature of advanced knowledge–based text information systems. The representation structures resulting from a proper text parse will also have to be susceptible to other transformational routines rendering *supplementary* services, such as fact retrieval, passage retrieval (a kind of backup mechanism for possibly insufficient partial text analyses), text filtering, indexing, etc. Text graphs are a plausible, yet experimentally validated candidate to serve as a unifying representation device under which those different applications could be organized in a coherent way.

6 Desiderata for Future Text Condensation Systems

1. Text graphs so far have been kept isolated from each other. What is needed is a *text graph pool* with uniform access mechanisms in order to facilitate text knowledge exploration on a significantly higher level of abstraction than the one available from pure, i.e., uncondensed text knowledge bases.

2. User interaction as well as text graph derivation were neither based on *planning* devices nor were explicit *user models* incorporated. Notions such as "relevance", "interestingness", "significance", "news", etc. are introduced into the system on a manual (interactive) basis. This constitutes a major shortcoming which hampers (actually, it precludes) proper representation of and reasoning about users' needs, expectations, dispositions, etc.

3. There exists no *formal* model of summarization. One should aim at calculi similar to relational algebra underlying the operation of data base systems or query logics for

question–answering systems. This should be feasible, since text summaries are transformed from knowledge structures whose formal semantics is well specified.

4. One should refrain from building special–purpose text understanders addressing message routing, summarization, knowledge extraction, etc. as exclusive modes of system functionality. Instead, we should turn to *general–purpose* designs for the entire process of text understanding and based on such a platform provide *specialized* transformation components capable of summarizing, question–answering, translating, message routing, etc. This amounts to saying that the language understanding (parsing) phase is entirely *decoupled* from the text knowledge transformation phase.

5. It seems reasonable to design text condensation in such a way that its model incorporate the constituent phases of text understanding, text condensation, and text/knowledge retrieval. Based on such a *coherent* design, it should be possible to determine criteria in order to assess the appropriateness, validity, adequacy, etc. of the text understanding and condensation phase. In other words, what constitutes a reasonable text parse?, which level of summarization is appropriate under what kinds of circumstances? which external factors influence the effectiveness of a summary (its "felicity conditions")?

6. Considering the last point once again, the question has to be raised what constitutes a reasonable *evaluation* platform for the assessment of summarization performance. Are we really to count single propositions in the original text and compare the result to the propositions contained in a condensate? What kinds of metrics apply to different degrees of condensation (e.g., relative to different types of abstracts)? One obvious starting point for deeper considerations would be to supplement general graph complexity measures by additional constraints that incorporate semantic properties of knowledge representation languages.

References

1Hahn, U.; **Reimer**, U. **[1984]**. Computing text constituency: an algorithmic approach to the generation of text graphs. In: C.J. van Rijsbergen (Ed.), *Research and Development in Information Retrieval. Proc. 3rc Joint BCS and ACM Symposium.* Cambridge, England, 2–6 July, 1984. Cambridge: Cambridge U.P., pp.343–368.

2Hahn, U.; **Reimer**, U. **[1986a]**. TOPIC essentials. *COLING '86: Proc. 11th Intl. Conf. on Computational Linguistics.* Bonn, 25–29 Aug. 1986, pp.497–503.

3Hahn, U.; **Reimer**, U. **[1986b]**. Semantic parsing and summarizing of technical texts in the

TOPIC system. In: R. Kuhlen (Ed.), *Informationslinguistik. Theoretische, experimentelle, curriculare und prognostische Aspekte einer informationswissenschaftlichen Teildisziplin.* Tuebingen: M. Niemeyer, pp.153–193.

4Hahn, U.; **Reimer**, U. **[1988]**. Automatic generation of hypertext knowledge bases. *Proc. Conf. on Office Information Systems*. Palo Alto, Cal., March 23–25, 1988. New York/NY: ACM, pp.182–188.

5Reimer, U.; **Hahn**, U. **[1988a]**. Text condensation as knowledge base abstraction. *Proc. 4th IEEE/AAAI Conf. on Artificial Intelligence Applications [CAIA]*. San Diego, Cal., March 14–18, 1988. Washington, D.C.: Computer Soc. Pr. of the IEEE, pp.338–344.

6Reimer, U.; **Hahn**, U. **[1988b]**. Knowledge–based text analysis in office environments: The text condensation system TOPIC. In: W. Lamersdorf (Ed.), *Office Knowledge: Representation, Management, and Utilization. Selected Full Papers Based on Contributions to the IFIP TC8/WG8.4 Intl. Workshop*. Toronto, Canada, 17–19 Aug. 1987. Amsterdam: North–Holland, pp.197–215.

7Reimer, U.; **Hahn**, U. **[1990]**. An overview of the text understanding system TOPIC. In: U. Schmitz, R. Schuetz, A. Kunz (Eds.), *Linguistic Approaches to Artificial Intelligence*. Frankfurt/M.: P. Lang, pp.305–320.

8Hahn, U. **[1987]**. A generalized word expert model of lexically distributed text parsing. In: B. du Boulay, D. Hogg, L. Steels (Eds.), *Advances in Artificial Intelligence – II. 7th European Conf. on Artificial Intelligence [ECAI–86]*. Brighton, U.K., July 20–25, 1986. Amsterdam etc.: North–Holland, pp.417–425.

9Hahn, U. **[1989]**. Making understanders out of parsers: Semantically driven parsing as a key concept for realistic text understanding applications. *International Journal of Intelligent Systems, 4* (3), pp.345–393.

10Hahn, U. **[1990a]**. Topic parsing: accounting for text macro structures in full–text analysis. *Information Processing & Management, 26* (1), pp.135–170.

11Hahn, U. **[1990b]**. *Lexikalisch verteiltes Text–Parsing. Eine objektorientierte Spezifikation eines Wortexpertensystems auf der Grundlage des Aktorenmodells*. Berlin: Springer.

12Hahn, U. **[1992]**. On text coherence parsing. *COLING '92: Proc. 15th Intl. Conf. on Computational Linguistics*. Nantes, 23–28 Aug. 1992, Vol. 1: Topical Papers, pp.25–31.

13Kuhlen, R. **[1984]**. *A knowledge–based text analysis system for the graphically supported production of cascaded text condensates*. University of Constance, Information Science Department *(Technical Report TOPIC–9/84)*.

14Hahn, U.; **Hammwoehner**, R.; **Reimer**, U.; **Thiel**, U. [1990]. Inhaltsorientierte Navigation in automatisch generierten Hypertext–Basen. In: P.A. Gloor, N.A. Streitz (Eds.), *Hypertext und Hypermedia – Von theoretischen Konzepten zur praktischen Anwendung*. Berlin etc.: Springer, 205–219.

15Hahn, U. **[1989]**. Dialogstrukturen in Gruppendiskussionen: Ein Modell fuer argumentative Verhandlungen mehrerer Agenten. In: D. Metzing (Ed.), *GWAI 89 – Proc. 13th German Workshop on Artificial Intelligence*. Eringerfeld, 18.–22. Sept. 1989. Berlin etc.: Springer, pp.409–420.

16Hahn, U. **[1992]**. Erklaerung als argumentativer Gruppendiskurs. In: H. Stoyan (Ed.), *Erklaerung im Gespraech – Erklaerung im Mensch–Maschine–Dialog*. Berlin etc.: Springer, pp.128–149.

17Hahn, U.; **Jarke**, M.; **Eherer**, S.; **Kreplin**, K. **[1991]**. CoAUTHOR: A hypermedia group authoring environment. In: J.M. Bowers , S.D. Benford (Eds.), *Studies in computer supported cooperative work: Theory, practice and design*. Amsterdam etc.: North–Holland, pp.79–100.

18Hahn, U.; **Jarke**, M.; **Rose**, Th. **[1991]**. Teamwork support in a knowledge–based information systems environment. *IEEE Transactions on Software Engineering, SE–17* (5), pp.467–482.

A Corpus–Based Approach to Domain–Specific Text Summarization: A Proposal

Ellen Riloff

Department of Computer Science
University of Massachusetts
Amherst, MA 01003
riloff@cs.umass.edu

1 Introduction

Summarizing texts effectively involves many different skills, but one of the most crucial is the ability to identify the most important information in a text. A good summary should include the most relevant aspects of a text but omit details and irrelevant information. However, it is often difficult to define what is important in the absence of a specific goal. In fact, different pieces of information will be relevant to different people depending on their individual interests and needs.

When we describe text summarization in these terms, the problem sounds remarkably like an information retrieval (IR) task. Information retrieval systems are concerned with identifying texts that contain information that is relevant to a user. A user presents the system with a specification of his or her information needs via a query or topic description. Ideally, the IR system will retrieve only the texts that contain information that is relevant to the query or topic. In general, text summarization does not require any topic description. But in some applications, it would make sense for a text summarization system to focus explicitly on a user's interests so it could tailor its summaries specifically to those interests. This scenario is the one that we will address; we will refer to this task as *domain–specific text summarization*. For our purposes, we will consider a query or topic description to be the same thing as a domain specification.

Text classification is an IR task that involves assigning one or more category labels to a text. We have developed several text classification algorithms that are based on a natural language processing (NLP) task called *information extraction*. In developing our text classification system, we have come to appreciate that information extraction, text classification, and text summarization share a lot of the same problems and goals. In particular, they all require the ability to extract relevant concepts without getting bogged down or misled by irrelevant information. Our text classification system uses a corpus–based approach to automatically identify relevant concepts that are highly correlated with a category. We will propose that this mechanism can also be used to automatically identify relevant portions of a text that should be included in a summary of the text.

In Section 2, we will briefly describe information extraction and explain how it relates to text summarization. In Section 3, we will describe a case–based text classification algorithm that we have developed which is built on top of an information extraction system. In Section 4, we suggest how our approach to text classification could be used to automatically summarize texts.

2 Information Extraction

In recent years, information extraction (IE) has emerged as a subtask of natural language processing that focusses on extracting domain–specific information from texts [Lehnert and Sundheim, 1991; MUC–3 Proceedings, 1991; MUC–4 Proceedings, 1992]. For example, an IE system in the domain of terrorism might extract the names of all perpetrators, victims, physical targets, and weapons that are mentioned in the context of a terrorist event.

In a broad sense, information extraction systems can be viewed as narrow–minded text summarization systems. On the one hand, they automatically skip over information that is not relevant to the domain. Since they retain only the domain–specific portions of the text, the output of an IE system can be thought of as a domain–specific summary of the text. However, since they retain virtually *all* of the information that is relevant to the domain, an IE system does not necessarily filter out much information. In the extreme case, a system may extract nearly every phrase that appears in a document.

Since they are designed to extract only domain–specific information, IE systems can be useful as a first pass filter to remove irrelevant information. However, within the domain, they do not discriminate between *important* information and details or peripheral information. Therefore an additional mechanism is needed to sift through the domain–specific parts of a text and generate a more discriminating summary.

3 Information Extraction as a Basis for Text Classification

We have developed several algorithms that use an underlying information extraction system to classify texts. The goal of our system is to classify texts as either *relevant* with respect to a given domain (i.e., the text contains domain–specific information) or *irrelevant* to the domain.

However, the system does not require an explicit set of domain guidelines or specifications from a user. Instead, the system exploits a "training corpus" of representative texts that a user or domain expert has manually classified as either relevant or irrelevant. We then use statistical techniques to sift through the corpus and automatically identify concepts that are reliable indicators of the domain. By requiring a training corpus rather than an explicit set of domain specifications, we ease

the burden on the user considerably. It is notoriously difficult for people to generate domain descriptions that are complete and accurate. But it is more straightforward for someone to skim a set of documents and pick out the ones that are of interest to them.

We have developed three different algorithms that use varying amounts of extracted information to classify texts [Riloff and Lehnert, 1992; Riloff, 1993; Riloff and Lehnert, 1994]. In the next section, we will describe one of these algorithms: a case–based text classification algorithm.

3.1 Case–Based Text Classification

The case–based text classification algorithm involves two phases: a training phase and a testing phase. In the training phase, we use the training corpus to generate a case base of natural language contexts. In the testing phase, we classify new texts by probing the case base for natural language contexts that are similar to the ones in the new texts.

During training, we use a conceptual sentence analyzer called CIRCUS [Lehnert, 1991] to extract information from the training texts. For each text, CIRCUS generates a set of instantiated *concept nodes*. A concept node is essentially a case frame that is activated by an individual word in a specific context and contains slots to extract specific types of information from the sentence. For example, in the domain of terrorism, a concept node called $damage–passive$ is activated by passive forms of the verb "damaged" and contains slots to extract targets, instruments, and perpetrators.

CIRCUS may generate zero, one, or several instantiated concept nodes for a sentence depending upon how many words in the sentence activate concept nodes. For each sentence, the case–based algorithm merges all of the concept nodes produced by the sentence into a single case structure. A text is therefore represented as a set of cases, one for each sentence that generated at least one concept node. For each slot filler, we save only the semantic features associated with the extracted information to generalize from the specific words in the text. Each case is stored in a case base which is used later to classify new texts. Figure 1 shows a sample sentence, the concept nodes produced by CIRCUS for the sentence, and the resulting case.

Sentence:

Two vehicles were destroyed and an unidentified office of the agriculture and livestock ministry was heavily damaged following the explosion of two bombs yesterday afternoon.

Concept nodes

$destruction–passive$ (triggered by destroyed)
target = `two vehicles`

$damage–passive$ (triggered by damaged)
target = `an unidentified office of the agriculture and livestock ministry`

$weapon–bomb$ (triggered by bombs)

Case

Signatures:
(<destroyed, $destruction–passive$>, <damaged, $damage–passive$>, <bombs, $weapon–bomb$>)
Perpetrators: nil
Victims: nil
Targets: (GOVT–OFFICE–OR–RESIDENCE TRANSPORT–VEHICLE)
Instruments: (BOMB)

Figure 1: A sample sentence, concept nodes, and resulting case

To classify a new text, we apply CIRCUS to each sentence and merge the concept nodes into a case. Then we determine whether any of the cases are "relevant". If the text contains a "relevant" case then we classify the text as relevant. Otherwise, we classify the text as irrelevant. A case is judged to be relevant if three conditions are satisfied:

Condition 1: The case contains a strong *relevancy index*.
Condition 2: The case does not contain any "bad" signatures.
Condition 3: The case does not contain any "bad" slot fillers.

The first condition is the heart of the algorithm. Given a new case, the most obvious approach would be to look for exact matches in the case base. However, cases typically contain many pieces of information so it is unlikely that we will find many exact matches in the case base. Instead, we probe the case base with a subset of the features in the case which we call a *relevancy index*. If the index retrieves mostly relevant cases then we assume that this is not a coincidence and the cases must all have something in common that makes them relevant. Since the cases all share the information represented by the index, we assume that the new case (which also shares the index) is relevant as well.

We define a relevancy index as a triple of the form: (signature, slot filler, case outline). A signature is a pair consisting of a word and a concept node that it triggers. Each signature represents a set of linguistic expressions recognized by the word/concept node pair. For example, the signature <dead, $found–dead–passive$> represents expressions such as "was found dead", "were found dead", etc. A slot filler is a pair consisting of a concept node slot and a semantic feature representing the information extracted by the slot. For example, "two vehicles" in the sentence in Figure 1 is represented by the pair (target, TRANSPORT-VEHICLE). Finally, the case outline is a list containing of the names of the slots that extracted information. For example, the case in Figure 1 would yield the outline: (targets instruments) because only these two slots are filled.

Since a case often contains multiple signatures and slot fillers, there may be many different relevancy indices for a case. We generate all possible relevancy indices exhaustively and, for each one, probe the case base to retrieve cases that share the index. If a high percentage of the retrieved cases came from relevant texts, then we assume that the index is a "strong" relevancy index.

Finally, if Condition 1 is satisfied then we check Conditions 2 and 3. These conditions determine whether any of the signatures or slot fillers in the case are **poorly** correlated with relevance. If so, then we do not classify the case as relevant. We have tested the case–based text classification algorithm extensively for the domain of terrorism using the MUC-4 corpus. A more detailed description of the algorithm and experimental results can be found in [Riloff, 1993; Riloff and Lehnert, 1994].

4 Relevancy Indices as a Basis for Text Summarization

The case–based algorithm represents a text as a set of cases, where each case contains the domain– specific information extracted from a single sentence. To classify a text, the algorithm identifies the most "relevant" cases, i.e., the cases that contain information that is highly correlated with the domain. Typically, only a subset of the cases will be deemed relevant.

For the purpose of text summarization, the relevant cases could be used to generate a summary of the text. These cases represent the portions of the text that are most strongly associated with the domain, based on statistics from the training corpus. Cases containing only general information are unlikely to be highly correlated with the domain because similar cases will be found in irrelevant as well as relevant texts in the training corpus. Cases containing only details and peripheral information are also unlikely to be highly correlated with the domain because there will be only a few, if any, similar cases in the case base. Therefore, the statistical nature of the algorithm will identify cases that contain key domain–specific information.

The relevant cases could be used in several ways. First, each case must be mapped back to its source sentence in the text. The simplest approach would be to use the entire case. That is, the summary would include the domain–specific portions of the sentence that are represented in the case. A more discriminating approach would include only the portions of the sentence that correspond to strong relevancy indices in the case. This method would produce smaller summaries that included only the information that was most highly associated with the domain.

As an example, consider the following text:

> *SAN SALVADOR, 13 JAN 90 (AFP) – [TEXT] The Salvadoran opposition National Revolutionary Movement (MNR) today demanded that Guatemalan President Vinicio Cerezo "thoroughly" investigate the* **assassination of MNR leader Hector Oqueli and of Guatemalan attomey Gilda Flores**.
>
> *During a news conference, the MNR executive committee condemned the "abominable" crime perpetrated on 12 January in Guatemala by armed men wearing civilian clothes. The MNR also demanded that the Salvadoran government guarantee the security of MNR Secretary General Guillermo Ungo, who will arrive in El Salvador within the next few hours to attend Oqueli's funeral, which is scheduled for 15 January.*
>
> **Oqueli, 45, and Gilda Flores, a member of the Democratic Socialist Party (PSD) of Guatemala were kidnapped** *while they were on their way to La Aurora airport in Guatemala City. Their bodies were found later at a place near the Salvadoran border.*
>
> *"The people responsible for this unjustifiable assassination are the eternal enemies of liberty, peace, and democracy," the MNR states. The MNR, along with the Social Democratic Party and the Social Christian People's Movement [MPSC], make up the Democratic Convergence Coalition.*
>
> *The MPSC also issued a communique deploring the* **assassination of Oqueli** *and urging the international community to condemn this action.*

The highlighted words correspond to information found in the cases for this text that are likely to be judged relevant. A summary based on this information might read: *"MNR leader Hector Oqueli and Guatemalan attorney Gilda Flores, a member of the Democratic Socialist Party (PSD) of Guatemala, were kidnapped and assassinated."*. This summary captures the essential information concerning the terrorist event. Note that the text contains additional information about the incident that our information extraction system would extract but which is not highlighted, e.g., that the crime was perpetrated on January 12 in Guatemala by armed men wearing civilian clothes.

The user could also control the length of the summary by adjusting the threshold values (1). When the threshold settings are high, the algorithm is conservative about classifying cases as relevant. High threshold settings would therefore produce smaller and more discriminating summaries. When the threshold values are low, the algorithm is more liberal about classifying cases as relevant so the resulting summaries would be longer.

5 Conclusions

We believe that domain–specific text summarization is a useful task in many real–world situations. While information extraction techniques can produce domain–specific summaries, additional mechanisms are necessary to filter through potentially large amounts of domain–specific information to generate concise summaries. We have proposed that a corpus–based approach to text classification could also be used to support text summarization. By exploiting a training corpus to identify key domain–specific concepts, the system automatically identifies the most relevant portions of a text that would be useful in a summary. We believe that this corpus–based approach to text summarization is a promising avenue for future research.

Footnote

(1)

The algorithm uses several thresholds to judge whether something is "highly" or "poorly" correlated with relevance.

References

Lehnert, W. G. and Sundheim, B. 1991.

A Performance Evaluation of Text Analysis Technologies. *AI Magazine* 12(3):81–94.

Lehnert, W. 1991.

Symbolic/Subsymbolic Sentence Analysis: Exploiting the Best of Two Worlds. In Barnden, J. and Pollack, J., editors 1991, *Advances in Connectionist and Neural Computation Theory, Vol. 1*. Ablex Publishers, Norwood, NJ. 135–164.

MUC–3 1991

Proceedings of the Third Message Understanding Conference (MUC–3). Morgan Kaufmann, San Mateo, CA.

MUC–4 1992

Proceedings of the Fourth Message Understanding Conference (MUC–4). Morgan Kaufmann, San Mateo, CA.

Riloff, E. and Lehnert, W. 1992.

Classifying Texts Using Relevancy Signatures. In *Proceedings of the Tenth National*

Conference on Artificial Intelligence. AAAI Press/The MIT Press. 329–334.

Riloff, E. and Lehnert, W. 1994.

Information Extraction as a Basis for Text Classification. *ACM Transactions on Information Systems*, Special Issue on Text Categorization (in press).

Riloff, E. 1993.

Using Cases to Represent Context for Text Classification. In *Proceedings of the Second International Conference on Information and Knowledge Management (CIKM–93)*. 105–113.

Automatic Summarizing – implementation and systems II

Rapporteur: Ralph Weischedel

Lisa F. Rau:

Domain-Independent Summarization of News

The first paper, presented by Lisa Rau, led to much discussion and interest. Because of copyright constraints, a commercial, automatic abstracting service for NEXIS was constrained to use a subset of the original source text; that is, deletion of text was the only admissible operation in producing an abstract.

Two strategies were evaluated:

- Deleting all sentences after the first few sentences (the number of sentences was limited by the maximum allowable abstract length).
- Deleting all sentences not exceeding criteria on importance (measured via tf * idf statistics, position in source article, absence of anaphora, etc.)

In the evaluation, the first strategy was judged to produce acceptable abstracts 90% of the time, compared to 75% for the more sophisticated strategy.

Discussion focussed on details of the evaluation methodology, for example, that the two evaluators may have had an interest in the first, simple strategy succeeding, that the evaluators could easily determine which strategy produced each abstract, and that the evaluators do not necessarily represent users or user needs. Nevertheless, Lisa pointed out that her conclusion was that those factors do not completely explain the evaluation result.

Another point emerging from the discussion was that the result should not surprise us too much. Since journalists are encouraged to write articles that enable readers to stop reading an article at any point, the challenge for automatic abstracting is to find effective solutions for documents that do not have that property.

Woojin Paik:

Chronological Information Extraction System (CIES)

The second paper proposed an approach to collecting facts about a named entitiy, e.g., a person, across many documents, and an approach to facilitating browsing documents clustered around events involving the individual.

Three points emerged from discussion:

- First, the clustering algorithm runs independent of the semantic network, and serves as a means of exploring the semantic network.
- Second, name changes (e.g. ˆMargaret Thatcher˜ vs. ˆLady Thatcher˜) are caught to the extent that at least one document uses the two forms as aliases.
- Third, the name categorization algorithm does not learn from experience, but does robustly identify names previously unseen.

Domain–Independent Summarization of News

Lisa F. Rau
Ron Brandow

Information Technology Laboratory
GE Corporate Research and Development
Schenectady, NY 12301

Karl Mitze

Data Enhancements
Mead Data Central
Miamisburg, OH

1 Introduction

In a recently completed system (1), we implemented and tested a prototype system to perform domain–independent summarization of news called ANES (Automatic News Extraction System). This draft extended abstract details the goals of the project, the implementation and the evaluation of the results.

1.1 General and Specific Research Goals

This effort was an experiment in the technical feasibility of automatically summarizing news. The primary objective was to ascertain and implement techniques that could address the problem of publication–independent summary generation; the hope was to produce a system of sufficient accuracy to be close to operational.

The prototype effort was geared towards accomplishing a number of objectives in addition to the demonstration of the functionality. In particular:

Source–independent:
The prototype was designed to be publication–source–independent to reduce the software maintenance costs. This objective is also important to ensure that the summary generation

system performs adequately for a wide variety of sources to maximize the potential utility of the software.

Extensible:

The system was designed so that publication–specific or product–specific summaries could be generated. The base configuration operates in a publication–independent mode, but contains the infrastructure to add special conditions to tailor summaries to either particular types of articles or particular types of information requests.

Flexible:

The system was designed so that the user could vary parameters to change the system's performance, such as the length and type of summary generated.

Copyright Compliance:

Only deletion of text was performed, while retaining grammaticality to avoid any copyright restriction violations or other legal issues involving manipulation of the source material while retaining the existing copyright.

The next sections describe the implementation and the evaluation of the results.

2 Prototype Description

The ANES system was designed and implemented to contain core technologies that could be applied outside the particular application of extracting news, and be easily reconfigured to alter performance along a number of different dimensions. In addition, the system was designed to require no maintenance upon the addition or change of publications going through the system, although periodic retraining is desirable. ANES is divided into two major sections:

Reader:

The Reader section reads in the input text, converting it internally into tokens, sentences and paragraphs. Additionally, the Reader section totals the word occurrences for the document and calculates word weights.

Extractor:

The Extractor generates the summary (from the internal representation). It determines document extractability, performs sentence weighting and determines the particular sentences to be included in the summary.

The process of summary generation has five major constituents.

2. 1 Training

ANES utilizes individual word statistics to determine the sentences of interest. To properly determine words of interest, ANES must be cognizant of the expected frequencies of words in a sizable representative corpus. During training, expected frequencies of words in a training corpus are generated; it is the user's responsibility to ensure that the training corpus used is sizeable and fairly representative of the types of documents from which ANES will be expected to generate summaries.

Training is used to determine the typical frequency of occurrence of words averaged across represented publications. A representative collection of documents is input, and corpus, document and word statistics are output. This information is used by the Reader section to generate word weights, which in turn are used to determine signature words. ANES training is performed off–line; i.e., training is performed as a batch process.

2.2 TF*IDF

ANES determines sentence selection by segregating out a list of signature words. Signature word selection was performed using *tf*idf* (term frequency times inverted document frequency).

For each document, the frequency of the words in the document is compared to the frequency of the words in the training set. For each word in the document, ANES remembers the word and its number of occurrences. Each word is given a weight which is based upon the number of times it appears in the document with respect to the number of times it appears in the training corpus. Words which appear more often in the document than they do in the training data set receive a higher weight.

Using *tf*idf*, words whose weight suggest that they could convey topical or other important information are segregated into a list of signature words. To this list, the headline words which are infrequent and are not already represented by a signature word are added. This extended list of signature words represents words that are relatively unique to the document and thus are likely to indicate topic or convey other important information.

2.3 Sentence Weighting

Sentence weighting is used to determine which sentences in a document contain information relating to the main ideas in the document. The sentence weights are based upon the weights of individual signature words (which were chosen using the TF*IDF process). The sentence weights are then utilized by the sentence selection function to select the extract sentences.

The sentence weighting function receives, as input, the document sentences, and the signature words. The weight of the sentence is the sum of the weights of the signature words present in the sentence.

Sentences which contain signature words are grouped into sequences of sentences based upon adjacency. To enhance readability, single sentences which contain no signature words but separate sentence sequences containing signature words are also included. Likewise, the first one or two sentences of a paragraph are also added when the second or third sentence, respectively, of the paragraph contains signature words.

The validity of the sentence weights depends heavily upon the accuracy of determining signature words; any improvements in determining signature words will translate into better sentence selection.

2.4 Sentence Selection

Sentence selection chooses the sentences that will become the final document extract. For each document, a set of sentences is chosen based on a number of factors, such as the presence of signature words in the sentence, its location in the document, the presence of words signaling anaphora, the target length of the extract and the type of extract to be generated. Additionally, ANES allows sentence selection to be done either by sentence weight or by sentence location.

Sentence selection determines the final extract sentences utilizing the sentences' weights calculated as described above and the desired extract length (sentence selection by weight) or the sentences' location in the document and the desired extract length (sentence selection by position). The generated document extract length is within +/– ten words of the desired word count.

Sentence selection makes very rudimentary use of the structure of the document; e.g., sentence selection does not ensure that some sentences are taken from every subsection in a multisection document. The use of structure information and the forcing of sentences from every subsection may improve the quality of abstracts from multisection documents.

3 Prototype Testing and Results

A 11.1 Megabyte sample of text consisting of 20,849 stories from 43 sources encompassing 41 distinct publications was used for development testing. This material, referred to as the "training set" was used to test the accuracy of the system and provide a sample of representative documents to ensure that the extraction software could handle a variety of document types.

Another data sample, 517,544 bytes, consisting of 95 documents, referred to as the "regression set", was also used. This data set was to be used for regression analysis; a constant data set to track the improvement in the system from week to week.

After the development phase was completed, ANES software was formally tested. The final performance rating was 74.4%, in close agreement with the final informal development phase testing performance rating. Additionally, 1.2% of the stories were correctly marked as unextractable (consisting of short unrelated material for example), while 1.6% of the stories were incorrectly marked as unextractable.

The test of ANES involved generating document extracts of 60, 150 and 250 words for a test set of 250 documents (a total of 750 document extracts). (The next page contains the preliminary evaluation results of comparing the acceptability of the ANES summaries to the MDC summaries).

The summaries output by the system were compared to those output by a Mead Data Central product called *Searchable Lead*. This system outputs sentences in order until the target summary length is achieved. (The results of these summaries appear in the Appendix A).

These results indicate that simply taking the first portion of news stories results in a superior summary than our statistical/heuristic system. Based on this evaluation, our future plans are to weight the sentences appearing in the beginning of the stories more heavily, and focus our efforts on detecting those cases where taking the first portion of the stories does not yield a good summary.

Although we wish our summaries to be perfect, most summaries judged less–than–perfect *would not be detectable as such to a user*. This is because a user is not considering the question "is this a good summary of the full text?". The user assumes this to be the case, and might only doubt the assumption when confronted with a summary that completely misses the point or is not comprehensible. This is a critical issue, and we recommend that end user acceptance testing be instituted to verify that the basic functionality of a news extraction system system like ANES is desirable and what level of performance is acceptable.

Achieving higher acceptability is primarily limited by the constraint that ANES be publication–independent, although somewhat higher accuracies could be achieved by continuing refinements within the existing framework. The publication–independent limitation forced the implementation to be intelligent enough to automatically identify certain types of documents, for example, question–answer sessions, speeches, op–ed pieces, tables, and embedded lists of various kinds, as opposed to hard–coding the special conditions that identify these documents.

4 Conclusions

ANES was an experiment in automatically summarizing news using a combination of statistical and heuristic methods. The most surprising result of the experiment was the adequacy of producing summaries consisting only of the first sentences in a story. Although the evaluators were not task–driven (that is, their judgements of adequacy were not with respect to task, but application–independent), we suspect that this type of summary is suitable for many purposes.

In spite of the difference in performance between the two methods, we are encouraged by the acceptability that was achieved with the statistical/heuristic method. This method allows for variable length summarization, as well as goal–directed summaries. A goal–directed summary is a summary that is particularized to a specified information need or goal a user is seeking; something the *Searchable Lead* system cannot do. In the future, this type of function should allow for highly individualized summaries to be maximally responsive to a user's information need.

Footnote

(1)

This paper was prepared while Lisa Rau was on an NSF Visiting Professorship for Women grant (NSF GER–9350134), hosted by the Computer and Information Sciences Department at the University of Pennsylvania.

Chronological Information Extraction System (CIES)

Woojin Paik

Syracuse University

Introduction

The research objective, which is described in this paper, is to provide rich contextual information about important entities discussed in newspaper texts. The current research is designed to take advantage of news reporters' practice of summarizing past events as background of the current story. In the news text schema (van Dijk, 1988), three discourse categories fall under the background aspect of the news story. They are *CIRCUMSTANCES*; *PREVIOUS EVENTS*; and *HISTORY*. We are currently focusing on the extraction of specific past events in which the proper names mentioned in the *MAIN EVENT* were involved. Text segments, which discuss these past events, are coded as *PREVIOUS EVENTS* or *HISTORY* by the Text Structurer of DR-LINK (Liddy et al, 1993). For example, a *HISTORY* segment referring to a person's purchase of an electronics company a few years ago, is important information for the system to extract, if the same person's purchase of a computer company is the *MAIN EVENT* in the current story. In addition, more general background information such as the age of a person or the location of the company is also extracted. This general information is coded as *CIRCUMSTANCES* by the Text Structurer. Text segments which are coded as *CIRCUMSTANCES* provide a description of the current context in which the proper name participates.

Important entities in news stories are often proper names. Proper names usually keep their original form or a recognizable variant while they are of information value. Thus, tracking one proper name across years worth of news stories is possible. Furthermore, background information of proper names is frequently expressed in predictable ways in newspaper texts. Thus, we are currently focusing on three types of linguistic construction which often provide background information on proper names. These constructions namely, appositive noun phrases, relative clauses, and copula sentences, are processed to extract relations between a proper name and pieces of background information. Relations define the nature of interaction, dependency, influence, or simple co–occurrence that hold between concepts. The relations allow us to create concept–relation–concept triples by treating proper names and pieces of background information as concepts. The concept–relation–concept triples are merged to form a semantic network.

We will add a browsing capability so that users of the system can explore the semantic network at their own initiative. We believe the browsing capability will be helpful for information seekers who

are not sure how to formulate clear search objectives nor have a sense of what information is available. We also believe the chronological information about proper names can be directly fed into database systems for fact retrieval.

The semantic network provides detailed information on proper names. One concept is usually linked to many concepts via many relations in a semantic network. Thus, the user is constantly presented with a large amount of information simultaneously. In addition, users are allowed to traverse the network unboundedly in the browsing mode. This browsing feature will maximize the possibility of finding all detailed information about a proper name. However, the semantic network is not intended to assist users to formulate a view of the general subject domain of a particular proper name. To provide this information, we are currently experimenting with a technique which classifies all stories which bear a common proper name. The classification of stories will present to the users a synopsis of the events in which the proper name was involved. For example, stories about H. Ross Perot, who is a Texas billionaire, might be categorized into two groups. One group of stories will be about business and the other will be about politics. The classification is accomplished by clustering a set of stories based on the subject content vector representation of each story (Liddy et al, 1992).

System Description

CIES is currently being developed as an extension of the DR-LINK System (Liddy et al, in press). DR-LINK is an information retrieval system which processes and represents texts at the lexical, syntactic, semantic, and discourse levels.

CIES executes the following four steps on newspaper texts to build a browsable semantic network in which proper names which occur in the texts are linked to their background information through meaningful relations.

Step 1. Identifying candidate texts which indicate the background information of a proper name.

- If an apposition belongs to the apposition proper category (Meyer, 1992) then the apposition is considered a candidate. An apposition proper consists of at least one coreferential noun phrase, which is a proper noun.

- If a relative clause modifies a proper noun then the clause is also considered as a candidate.

- If the subject of a copula sentence is a proper noun then the sentence is considered as a candidate.

- Thus, CIES relies on part of speech information and the phrase structure of sentences to determine candidates. The part of speech information is provided by POST (Meteer et al, 1991), a probabilistic part of speech tagger at the preprocessing stage of DR-LINK. Phrases and clauses are SGML marked by a finite state automata phrase/clause bracketter.

Step 2. Extracting relations between proper names and their background information.

- CIES relies on two different approaches for extracting relations. The first approach deals with appositions and predicts a relation between a coreferential proper name and the proper name in the appositive noun phrase based on both proper names' semantic category. The semantic categories of the proper nouns in texts are assigned by the Proper Noun Interpreter of DR-LINK (Paik et al, 1993), which classifies each proper noun in text as one of thirty–seven categories and produces a canonical representation of each proper name. Some examples of the proper name categories are person, company, organization, government, disease, and city. An example of a rule is that if there is a name of an organization in the appositive noun phrase and the coreferential proper noun is a person name then the relationship between the company and the person is most likely affiliation.

- The second approach extracts relations using Relation Revealing Formula (RRF) (Liddy, 1989). RRFs can be regarded as a sublanguage approach (Sager et al, 1987) to analyzing texts. Sublanguage theory suggests that any type of text that is used for a common purpose within a group of individuals will develop characteristic syntax and semantics. RRFs used in CIES are developed by an intellectual analysis of high–frequency phrases with semantic similarities, but expressed in a variety of phrases in newspaper texts. For example, the following RRF delineates the *ROLE* relation between the coreferential proper noun and the animate noun inside of the appositive noun phrase.

PERSON_PN, COMPANY_PN's {JJ/JJR/JJS} **ANIMATE_NN** ,...

PN: proper noun, JJ: adjective, JJR: comparative adjective,
JJS: superlative adjective, NN: singular
noun, {}:zero or more occurrences, /:optionality

Figure 1. ROLE RRF #002

Step 3. Building concept–relation–concept (CRC) triples.

- CRC triples are usually generated by linking a proper name of interest and its background information through an extracted relation from the previous step. It is possible to have more than one piece of background information from one candidate text be linked to the proper name of interest via more than one relation (Fig. 2).

- The following figures show candidate texts, which were used to extract relations, and CRC triples, which were generated by CIES. All words in the candidate texts are tagged with the Penn Treebank Part–of–Speech (POS) tag sets (Santorini, 1991). A bar, '|' is used to segregate a word from a POS tag. Each proper noun has an index number attached after the POS tag. Index numbers point back to the canonical form of each proper noun, which is stored in a separate field at the beginning of each document. Pseudo–SGML tags are used to bracket different types of phrases and clauses. For example, '<APP>' represents the beginning of an appositive noun phrase; '<N>' represents the beginning of a noun phrase; and '<RC>' represents the beginning of a relative clause. Some phrase/clause SGML tags have been removed from the original texts to enhance the readability of the texts. In the output CRC triples, proper names are shown in the canonical form along with semantic category information. Words inside of parentheses represent the names of the extracted relations. Arrows represent the direction of the relation. A relation links a concept which is located directly above the relation with the concept which is located to the right of the relation.

```
candidate:

Mr._Perot|NP|7 ,|, <APP> GM|NP|4 's|POS <N> largest|JJS individual|JJ shareholder| NN <N>
</APP> ,|, ...

output CRC triples:

H._Ross_Perot|Person
-> (AFFILIATION) ->     General_Motors|Company
-> (ROLE) ->            "largest individual shareholder"
```

Figure 2. appositive noun phrase based CRC triples (from WSJ861201-0119)

- Two CRC triples in Fig. 2 capture the circumstantial information about H. Ross Perot, which is a person proper name. The *AFFILIATION* relation links H. Ross Perot and General Motors, a company with which Perot is associated. The *ROLE* relation between

H. Ross Perot and the noun phrase, which is headed by an animate noun, 'shareholder', reveals how Perot is related to General Motors. The RRF which is used to extract *ROLE* relation is shown in the Fig. 1.

```
candidate:

H._Ross_Perot|NP|4 ,|, <RC> who|WP resigned|VBD Monday|NP|5 as|IN <N> the|DT head|NN </N>
of|IN GM|NP|6 's|POS Electronic_Data_Systems_Corp.|NP|7 and|CC as|IN <N> a|DT GM|NP|6
director|NN </N> </RC>,|, ...

output CRC triples:

H._Ross_Perot|Person
-> (AGNT) ->    resign
                -> (TIME) ->    Monday
                -> (ROLE) ->    "the head"
                -> (AFFILIATION)->Electronic_Data_Systems_CORP|Company
                                -> (PARENT) -> General_Motors|Company
                -> (ROLE) ->    director
                -> (AFFILIATION)->General_Motors|Company
```

Figure 3. relative clause based CRC triples (from WSJ861202-0030)

- CRC triples in Fig. 3 capture the information about an event in which Perot was involved. Perot is linked to the verb, 'resign', through *AGENT* which is a case relation. The verb describes the action of Perot in the event. In addition, the typical personal background information, which is bounded by the *ROLE* and *AFFILIATION* relations, is linked to 'resign'. This verb serves as the reason why the relations *ROLE* and *AFFILIATION* between Perot and the background information become ineffective. The *TIME* relation between 'resign' and the temporal information captures the point in time when the action occurred.

```
candidate:

Mr._Perot|NP|4 is|VBZ extremely|RB close|RB to|TO <N> his|PP$ son|NN </N> .|.

output CRC triples:

H._Ross_Perot|Person
-> (STATE) -> "extremely close"
           -> (CHILD) -> "his son"
```

Figure 4. Copula Sentence based CRC triples (from WSJ861202-0014)

- CRC triples in Fig. 4 capture circumstantial information about Perot. The matrix verb, 'is' signals *STATE* relation between Perot and the adverbial phrase, 'extremely close'. The *CHILD* relation is revealed by the semantic feature of the animate noun, 'son'.

Step 4. Merging CRC triples with common proper names.

- The proper name's originating document identification and its date of publication are added to the CRC triples through *source* and *date* relations.

- CRC triples are then merged if they share a common proper name. The merging operation follows principles described as the *rule of join* and the *rule of simplification* in Sowa (1984). The *rule of join* is defined as deleting all common concepts except one, then linking relations that had been linked to the deleted concepts, to the undeleted concept. The *rule of simplification* is defined as deleting all common relations, which are linked to the same concept, except one. A semantic network, which consists of concepts and relations, is generated at the end of this step.

Following is an additional function of `CIES` which can be called while the users browse the semantic network created in step 4.

Step 5. Dynamic clustering of source documents for each proper name.

- A set of documents, which mention the same proper name, are clustered using Ward's (1963) agglomerative clustering algorithm on demand. The tree–like representation of clusters is also browsable by the users. Each document in the clusters is represented by the headline of the story. `CIES` clusters the Subject Field Code (SFC) vector representations of stories (Liddy et al, 1992). SFCs are categories from the subject classification scheme used in *Longman Dictionary of Contemporary English* (LDOCE) (1987). SFCs are available for senses of words in LDOCE. To create a SFC vector of a story:
    - each word in the story is tagged with the automatically sense disambiguated SFC; and
    - the within–story SFCs are summed and normalized.

  SFC vectors represent texts at a more abstract conceptual level than the lexical level in natural language text.

Generally speaking, an event **E**, which is the main focus of a story at that point in time (**T**) when it is reported will become background of a new event **E'** at a future time (**T+1**) if there is enough relatedness between two events **E** and **E'**. Event **E** will then be presented in a summarized form when it is reported as a previous event of the current event **E'** in a later story. In CIES, two events are considered related if they share a common proper name. Thus, CIES' current capability of extracting background information about one proper name over a period of time can be thought of as collecting summaries of events in which the proper name of interest was involved.

The implementation of CIES has just begun. Thus, there are no empirical results with which to evaluate the system.

References

Liddy, E.D. (1989).
Explorable vocabularies. *Proceedings of the first International Lexical Acquisition Workshop*, sponsored by AAAI in conjunction with IJCAI–89, Detroit.

Liddy, E.D., Paik, W. & Woelfel, J. (1992).
Use of subject field codes from a machine–readable dictionary for automatic classification of documents. *Proceedings of 3rd ASIS Classification Research Workshop.*

Liddy, E.D., McVearry, K., Paik, W., Yu, E.S. & McKenna, M. (1993).
Development, implementation & Testing of a Discourse Model for Newspaper Texts. *Proceedings of the ARPA Workshop on Human Language Technology*. Princeton, NJ, March 21–24, 1993.

Longman Dictionary of Contemporary English (1987).
2nd Ed., Longman Group, Harlow, England.

Meteer, M., Schwartz, R. & Weischedel, R. (1991).
POST: Using probabilities in language processing. *Proceedings of the Twelfth International Conference on Artificial Intelligence*. Sydney, Australia.

Meyer, C.F. (1992).
Appositions in contemporary English. Cambridge, England: Cambridge University Press.

Paik, W., Liddy, E.D., Yu, E.S. & McKenna, M. (1993).
Categorizing and standardizing proper nouns for efficient information retrieval. *Proceedings of the ACL Workshop on Acquisition of Lexical Knowledge from Text*. Columbus, OH.

Santorini, B. (1991).
Part–of–Speech Tagging Guidelines for the Penn Treebank Project. *Technical Report.* University of Pennsylvania.

Sager, N., Friedman, C., & Lyman, M.S. (1987).
Medical Language Processing: Computer Management of Narrative Data. Reading, Mass.: Addison–Wesley.

Sowa, J. (1984).
Conceptual Structures: Information Processing in Mind and Machine. Reading, MA: Addison–Wesley.

Ward, J. (1963).
Hierarchical grouping to optimize an objection function. *Journal of the American Statistical Association.* 58, p. 237–254.

van Dijk, T. (1988).
News analysis: Case studies of international and national news in the press. Hillsdale, NJ: Lawrence Earlbaum Associates.

Automatic Summarizing – implementation and systems III

Rapporteur: Annely Rothkegel

Elizabeth D. Liddy:

Development and Implementation of a Discourse Model for Newspaper Texts

There are four main discussion points:

1. text structure:
Is it – in principle – possible to construct a global text structure from particular text components? Are there some general criteria for determining what combination of parameters are relevant?

2. text type:
How independent of the text type a text model can be created?

3. information extraction:
What relationship does exist between text structure and options of information extracting?

4. Coding problems:
What kind of problems do arise with respect to the training of the people who do the coding work?

Mark T. Maybury:

Automated Event Summarization Techniques

1. General discussion:
What might be a (perfect) summary?
What kind of inferences should be provided/prevented?
Is it possible to define some types of summaries?
Could it be possible to determine some prototypes?

2. Summarizing techniques and implementation:
For improving implementation it would be helpful to consider the results of the research in this field of the last 15 years!

Development and Implementation of a Discourse Model for Newspaper Texts

Elizabeth DuRoss Liddy

School of Information Studies,
Syracuse University
Syracuse, New York, USA
liddy@mailbox.syr.edu

Introduction

In this paper, we will focus on the development, implementation, and evolution of a discourse model which is used to computationally instantiate a discourse structure in individual texts. This discourse model was developed for use in a Text Structuring module that recognizes discourse–level structure within a large–scale information retrieval system, DR–LINK (Liddy & Myaeng, 1993). The Text Structurer produces an enriched representation of each document by computationally decomposing it into smaller, conceptually labelled components. This delineation of the discourse–level organization of each document's contents facilitates retrieval of those documents which convey the appropriate discourse semantics that are responsive to the user's query.

The recognition of the existence of text–type models derives from research in discourse linguistics which has shown that writers who repeatedly produce texts of a particular type are influenced by the schema of that text–type and, when writing, consider not only the specific content they wish to convey but also what the usual structure is for that type of text based on the purpose it is intended to serve (Jones, 1983). As a result, texts of a particular type evidence the schema that exists in the minds of those who produce the texts. These schema can be delineated, and as such provide models of their respective text–types which are of use in automatically structuring texts. A text schema explicates a discernible, predictable structure, the global schematic structure that is filled with different meaning in each particular example of that text–type (van Dijk, 1980). Among the text–types for which schemas or models have been developed are: folk–tales (Propp, 1958), newspaper articles (van Dijk, 1980), arguments (Cohen, 1987), historical journal articles (Tibbo, 1989), editorials (Alvarado, 1990), empirical abstracts (Liddy, 1991), and theoretical abstracts (Francis & Liddy, 1991).

Development of the News Text Schema

Our first effort towards including discourse–level semantics in the DR–LINK System was focused

on newspaper texts, taking as a starting point, the hierarchical newspaper text model proposed by van Dijk (1988) . Several iterations of human analysis and coding of a sample of randomly selected Wall Street Journal articles using the components from van Dijk's model motivated us to develop a revised News Schema which reorganized van Dijk's categories according to a more temporally–oriented perspective and added several new components. The resulting News Schema Components were: CIRCUMSTANCE, CONSEQUENCE, CREDENTIALS, DEFINITION, ERROR, EVALUATION, EXPECTATION, HISTORY, LEAD, MAIN EVENT, NO COMMENT, PREVIOUS EVENT, REFERENCES, and VERBAL REACTION.

The process of manually coding the training sample also served to suggest to us the different types of linguistic information which we implicitly relied on during our intellectual decomposing of texts. These intuitions were further explored by means of statistical analyses of the linguistic differences exhibited by text in the various components. These results were translated into computationally recognizable text characteristics for use by the Text Structurer to assign a single component label to each sentence. Briefly defined, the sources of evidence used in the first implementation of the Text Structurer were:

Lexical Clues – A set of one, two and three word phrases for each component, based on observed frequencies and distributions. Clues are words with sufficient occurrences, and a statistically skewed observed frequency of occurrence in a particular component. Not surprisingly, many of clues strongly suggest the semantic role or purpose of each component.

Order of Components – The tendency of components to occur in a particular, relative order determined by calculating across the coded training files.

Likelihood of Component Occurring – The observed frequency of each component in our coded sample set.

Tense Distribution – Some components, as might be expected by their name alone, tend to contain verbs of a particular tense more than verbs of other tenses.

Syntactic Sources – Two types of syntactic evidence:

1. typical sentence length as measured in average number of words per sentence for each component;
2. individual part–of–speech distribution based on the output of the part–of–speech tagging of each document, using POST, a part–of–speech tagger (Meteer et al, 1991).

Continuation Clues – Based on the conjunctive relations suggested in Halliday and Hasan's Cohesion Theory (1976), lexical clues which occur in a sentence–initial, or near sentence–initial

position, and which were observed in our coded sample data to predictably indicate either that the current sentence continues the same component as the prior sentence or that there is a change in the component.

These sources of evidence for instantiating a discourse–level model of the newspaper text–type were incorporated in the computational Text Structurer in our system, which evaluates each sentence of an input newspaper article against these evidence sources, comparing it to the known characteristics of each component of the text–type model, for the purpose of assigning a text–level label to each sentence.

The first computational implementation of the Text Structurer used the Dempster–Shafer Theory of Evidence Combination (Shafer, 1976) to coordinate information from the various evidence sources. In that implementation, each document is processed a sentence at a time, and each source of evidence assigns a value between 0 and 1 to indicate the degree of support that each evidence source provides to the belief that a sentence is functioning as a particular news–text component. The probability of each observed value for each piece of evidence for each component is calculated and is used as a belief in the Dempster–Shafer algorithm for evidence combination. Then, a simple supporting function for each component is computed and the component with the highest assigned belief value is selected as the correct component tag for that sentence.

Attribute Model of News Text

After completing the first implementation of this model, we moved to a new, attribute model of news–text structure. One factor which precipitated this move was the difficulty we encountered in manually coding some new training data. These difficulties appeared to be caused by our increased awareness of the multiple attributes or dimensions embedded in each of the component labels. For example, we realized that PREVIOUS EVENT was defined by a combination of particular values on the dimensions of Importance, Time, Completion, and Definiteness. Although each of the individual dimension values was shared by other components, PREVIOUS EVENT was a unique combination of dimension values. That is, although several components shared the same values on some dimensions, each component was distinguished from all other components by its value on at least one dimension. We felt that the more holistic tagging of sentences with component labels such as PREVIOUS EVENT, CIRCUMSTANCE, and LEAD had not adequately reflected these micro–level similarities and distinctions.

In addition, questions from members of the potential community of users of the structured output – questions such as: "Do the component labels indicate the status (e.g. journalist vs. participant in the news) of the views in the text?" or "Do the component labels indicate whether an event is ongoing or completed?" made us realize that the granularity of the components in the original model did not

explicitly indicate these facts, although they were implicit in the components' definitions which we had developed and relied on for manual coding. Therefore, we concluded that there was a dual need to:

1. capture and represent the basis of the commonality amongst some components, as well as;
2. make more distinct the uniquenesses which distinguished components.

In an attempt to accomplish these goals, we developed the Attribute Model of the news text in which pieces of text are evaluated for their specific value on each of eight dimensions or attributes: time of event, tense, importance, attribution, objectivity, definiteness, completion, and causality. Plus or minus values on these attributes were assigned to the text pieces without consideration of the component labels from the earlier model. At this point, we also began coding text at the clause rather than the sentence level, since we recognized that single sentences do contain multiple discourse–level components. These might be reflected in tense changes within a single sentence, or appositional statements of past events within a straight–forward reporting of a current news event.

After reviewing our recoding of the sample texts, we realized that the move to the Attribute Model had resulted in the loss of a very important function which had been performed by the earlier discourse–component labelling of sentences using the News Text Schema. That is, the recoded data seemed to prove the old adage that the whole is sometimes greater than the sum of its parts – that is, labelling a segment of text PREVIOUS EVENT had conveyed more than simply identifying that text segment's values on the eight dimensions. In other words, the discourse–component label conveys the role or function within the larger news–text model, information that is not conveyed by the Attribute Model coding. That is, discourse–level structured news articles based on the News–Text Schema convey a great deal of significant linguistic and pragmatic information that is not available without this discourse level analysis and processing.

Revised News Text Model

Although we recognized the superiority of the earlier News Text Schema over the newer Attribute Model, we did not want to lose the distinctions and similarities amongst text segments which we were able to recognize when using the eight dimensions of the Attribute Model. Therefore, we moved to a revision of our original News Text Schema, a refinement of the earlier components via addition of some of these distinguishing attributes to the earlier components. Operationally this was accomplished via the addition of sub–codes. For example, LEAD was sub–coded for its temporal aspect via the codes HISTORY, PREVIOUS, and FUTURE; CONSEQUENCE was sub–coded for PAST, PRESENT and FUTURE; EVALUATION was sub–coded for JOURNALIST to distinguish opinion which is not attributed to a source and therefore likely to be the journalist's view

from plain EVALUATION, which is an opinion attributed to a named source. In addition, PREVIOUS EVENT and HISTORY had sub–codings added for CONTINUOUS, and MAIN was sub–coded for FUTURE, as well as SECOND EVENT, and EXAMPLE.

Given this more complex News Text Schema, the original Text Structurer implementation which made use of eight sources of linguistic evidence did not appear reasonable for processing gigabytes of text for our DR–LINK Project. Based on an analysis of the automatically–structured documents produced by the first implementation, we measured how much each evidence source contributed to the system's ability to assign correct components. From that analysis, we determined that the more important evidence sources were lexical clues, tense data, and continuation clues. Therefore we reduced the number of evidence sources to these three. These evidence sources were evaluated heuristically by a combination of rules and lexicon. Frequency of occurrence of each component, sentence length, and distribution of parts of speech were dropped as evidence sources. Ordering information, which was ineffectively implemented as an evidence source in the first implementation, is currently being re–incorporated, as is a return to the Dempster–Shafer approach to evaluating and combining evidence.

The development of a leaner implementation in which only those evidence sources which contributed most significantly to the system's ability to correctly recognize pieces of text as particular components was used. The new implementation of the revised News–Text Schema instantiated a more precise model both in terms of the specificity of the model's components and the unit of text assigned a discourse component.

Conclusion

The process of developing, implementing and iteratively revising a discourse model for one text–type for use in the computational recognition of discourse–level structure in text is not yet finished. We have empirical results which indicate the News Text Schema's positive contribution to a text retrieval application by enabling DR–LINK to recognize documents which are relevant to a query on the basis of their discourse–level semantics as captured by the News Text Schema. We are currently engaged in efforts to both improve the current implementation as well as efforts to generalize the model.

Acknowledgements

The work reported herein would not have been possible without the fine efforts of Woojin Paik, Mary McKenna, Ken McVearry, Edmund Yu, and Carin Obad. The research was supported by ARPA's TIPSTER Initiative.

References

Alvarado, S. J. (1990).
Understanding editorial text: A computer model of argument comprehension. Kluwer Publishers.

Cohen, R. (1987).
Analyzing the structure of argumentative discourse. *Computational Linguistics*, 13, pp. 11–24.

Francis, H. & Liddy, E. D. (1991).
Structured representation of theoretical abstracts: Implications for user interface design. In Dillon, M. (Ed.). *Interfaces for information retrieval and online systems*. Greenwood Press.

Halliday, M. A. K. & Hasan, R. (1976).
Cohesion in English. London, Longmans.

Jones, L. B. (1983).
Pragmatic aspects of English text structure. Arlington, TX: Summer Institute of Linguistics.

Liddy, E. D. (1991).
The discourse–level structure of empirical abstracts: An exploratory study. *Information processing and management*, 27:1, pp. 55–81.

Liddy. E.D. & Myaeng, S. H. (1993).
DR–LINK: A system update for TREC–2. In Harman, D., (Ed.), *Proceedings of the second Text Retrieval Conference.*

Meteer, M., Schwartz, R. & Weischedel, R. (1991).
POST: Using probabilities in language processing. *Proceedings of the Twelfth International Conference on Artificial Intelligence*. Sydney, Australia.

Propp, V. (1958).
Morphology of the folk–tale. (L. Scott, trans.). Bloomington, Indian University Press. (Original work published 1919).

Shafer, G. (1976).
A mathematical theory of evidence. Princeton, NJ: Princeton University Press.

Tibbo, H. R. (1989).
Abstracts, online searching, and the humanities: An analysis of the structure and content of abstracts of historical discourse. PhD Dissertation, College of Library and Information Science.

van Dijk, T. (1980).
Macrostructures: An interdisciplinary study of global structures in discourse, interaction, and cognition. Hillsdale, NJ: Lawrence Earlbaum Associates.

van Dijk, T. (1988).
New analysis: Case studies of international and national news in the press. Hillsdale, NJ: Lawrence Earlbaum Associates.

Automated Event Summarization Techniques

Mark T. Maybury

Artificial Intelligence Center
Mail Stop K329
202 Burlington Road
Bedford, MA 01730
(617) 271–7230
maybury@mitre.org

Abstract

Automatically summarizing events from data or knowledge bases is a desirable capability for a number of application areas including report generation from databases (e.g., weather, financial, medical) and simulations (e.g., military, manufacturing, economic). While there have been several efforts to generate narratives from underlying structures, few have focused on event summarization. This extended abstract outlines tactics for selecting and presenting summaries of events. We discuss these tactics in the context of a system which generates summaries of events from an object–oriented battle simulator.

Topic Area:

Automated Summarization, Report Generation

1. Introduction

While there have been several investigations to generate narrative, fewer have looked explicitly at event summarization. For example, Goldman's BABEL (1975) within MARGIE produced stories by translating underlying conceptual dependency structures. Meehan's (1976) TALE–SPIN traced the goal stack of simulated agents to tell their story. In contrast to these story generators, Lehnert (1981) explicitly focused on summarizing events in stories from a text interpretation perspective. In an attempt to abstract away from the underlying events and states of a story, Lehnert suggested a number of "plot units" (e.g., problem resolution by intentional means, trade, and honored request) which were configurations of "positive" events, "negative" events, and "neutral" emotional states. Unfortunately, these structures simply do not provide sufficient constraints to guide content selection.

Guided by analysis of the discourse structure of narrative, many domain–specific event reporting patterns have been suggested. For example, Kukich's (1983) ANA produces stock market reports (regarding the trend and volume of the industrial average) by collecting semantically related messages concerning 10 specific market issues including "closing market status", "volume of trading", "mixed market" and "interesting fluctuations." A relatively simple "discourse organizer" groups and prunes messages (thus summarizing the content) which are then realized using a phrasal lexicon of just under 600 entries. Similarly, Kittredge et al. (1986) use a sublanguage approach in RAREAS to produce Canadian weather reports. The sublanguage includes knowledge about the saliency of content such as the fact that warnings preceded normal weather or that sentence groupings follow the order
`WINDS > CLOUD-COVER > PRECIPITATION > FOG&MIST> VISIBILITY.`

These sublanguage approaches can be contrasted with systems that represent explicit *text grammars* (domain specific schemata) such as Li et al.'s, (1986) system which produces two types of reports, a current status report and a discharge report, by accessing facts from a Stroke Consultant expert system. For example, the top level rule for a stroke case report is:

```
Case_Report -> Initial_Information + Medical_History +
Physical- Examination + Laboratory_Tests + Final_Diagnosis +
Outcome
```

The `Initial_Information` portion of this rule includes patient information, described by the rule:

```
Patient_Information -> Registration_Number + Age + Handedness
+ Race + Sex
```

The right–hand side of the `Patient_Information` rule consists of leaf nodes which refer to information in the stroke knowledge base. In addition to being domain dependent, this text grammar does not have an explicit indication of the relative importance of information in the report that could be exploited for summarization. While there have been contributions to domain–independent text schemata for descriptive texts (McKeown, 1985), like text schemata, the aforementioned text grammars and sublanguages are "compiled" plans that indicate *when* content should appear in a text but not *why*, so the ability to tailor summaries to individual contexts or users is not possible without extension (however, see Paris (1988)). While other researchers have taken more general approaches to narrative generation (e.g., Hovy's (1988) "structurer"), their explicit focus has not been on summarization. The next section outlines such techniques.

2. Tactics for Event Summarization

There are several classes of techniques for summarizing events from simulations or other event–oriented application systems or domains. Table 1 summarizes these tactics. These include exploiting the *saliency* of events (and their associated characteristics), abstracting more general events from collections of events, integrating related events, analyzing the types and numbers of links (i.e. semantic relations) between events, analyzing statistical distributions of events, and controlling the presentation of events. These techniques are not mutually exclusive, nor do we claim they are a complete set, rather they serve as a starting point for further refinement. We illustrate and discuss each of these in turn.

Implicit in the stock market, weather, and medical report generators described above is a notion of the importance of types of events or information. In the report generators discussed above, the domain–specific saliency of a particular entity is used to govern content selection and presentation order. Analogously, human–produced newspaper reports typically attempt to capture the key characteristics of some event (i.e., who, what, when, where, how, and why) in the leading or "topic" sentence or paragraph.

Saliency is a domain specific measure of the relative importance or prominence of an event, and can refer either to particular events, characteristics of events, or classes of events. For example, in a military simulation we have studied (Anken, 1989), domain experts indicate that events dealing with bombings or missile launches are, in general, more significant than movement events. And yet certain kinds of movements in certain contexts may be very important (e.g., those leading up to a bombing, or those resulting in strategic repositioning of assets). Just as classes of events can be specified as significant, so too particular events or patterns can be identified as significant (e.g., a bombing followed by an explosion; bombings of particular kinds of facilities). Importance is not only context dependent, but also in the eye of the beholder. For example, stereotypically, military operations staff are concerned with strike missions, close air support missions, and so on; logisticians care about refueling, resupply, and transportation missions; intelligence users care about enemy forces, their type, size, location, and activities. In addition to these broad classes of user interests, there are many other perspectives which could govern content selection (air vs. ground vs. sea; friendly vs. enemy). Also, for particular event classes, there may be particular semantic roles associated with types of events that domain specialists deem more salient than others (e.g., the target and weapon associated with a bombing event). Finally, event characteristics (e.g., their particular location, size, number of participants) may also indicate significance. While we have found event saliency to be effective for summarizing a battle simulation, it is domain specific and does not take advantage of context.

Another domain specific summarization technique, *abstraction*, takes a series of events and replaces them with a single event. Lehnert's (1981) "plot units" are a general case of this. A

domain specific example, in our battle simulation, is that a number of movement events followed by missiles firing, and an aborted mission event can be abstracted into a foiled attack event. That is, a number of subevents can be described by an overarching event. This abstraction can be accomplished either by using pattern–matching techniques or plan–recognition. While potentially very powerful, there can be significant additional cost to build the abstraction machinery, complicated by the need to validate more general events as being meaningful to the addressee. Moreover, leaving out information reduces detail, potentially increasing vagueness and ambiguity. The trick is to discover the "right" level of abstraction for the user.

Technique	Strengths	Weaknesses
1. Saliency (of events)	relatively easy to implement; intuitive	domain dependent; requires acquisition from experts; context independent
2. Saliency (of event attributes or semantic roles)	easy to implement; intuitive,	domain dependent; application must represent and reason about semantic roles and associated data; requires knowledge acquisition from experts
3. Abstraction	enables summarization across events; elides details	requires semantic hierarchy of events and more complex reasoning about abstraction; risk loss of precision and coherency by eliding details
4. Integration (semantic or linguistic)	requires less presentation space and time for same content (e.g., using adverbials to combine similar events)	requires identifying and integrating similar events into a single utterance; should be used in moderation
5. Link Analysis	takes advantage of type and number of semantic relations among events and states; domain independent	requires semantic relations in application; need mechanisms to capture event/state network; assumes isolated events unimportant
6. Statistical Analysis	easy to implement; domain independent; computationally inexpensive; easy to identify statistical outliers (in current set of events or historically)	relies on law of large numbers to measure event distributions, can overlook significant events embedded in large numbers of events of similar type
7. Presentation	context (time, space, topic) can be exploited to both reduce content and increase coherency; different media types (e.g., text vs. graphics, maps) may shorten elapsed time of presentation; different rhetorical techniques (e.g., define vs. describe) can elide information	requires representing and reasoning about context, media, and/or rhetoric; rhetorical tactics limited to structured prose

Table 1. Event Summarization Tactics

A related, less complex summarization technique, termed here *integration*, identifies similar events that can be unified into a single description. For example, two events that share a semantic agent, patient, instrument and so on, can be expressed in a single sentence. For example, if two missions are shot at the same time our battle report generator simply states "Site–A and Site–D *simultaneously* fired a missile at offensive counter air mission 102" using a temporal adverbial to relate the events. This technique both reduces the amount of space and time required to report the two events and increases the local cohesion of the text.

In contrast to selecting events based on semantic patterns, two domain independent event selection techniques are link analysis and statistical analysis. In the former case, the relative importance of particular events is determined by the amount and type of links between events. The assumption is that, for example, events that enable or cause many other events should be more significant than events that are isolated from all other events.

An even more basic method, particularly useful when analyzing a large pool of events, is to simply count the number of different types of events, the times they occur, and so on, and from this determine, statistically, which are most significant. The assumption is that in simulations or other event–oriented applications, events that occur frequently tend to be less significant than those that occur infrequently.

Consider, for example, Figure 1 which illustrates a histogram of occurrences of different types of events from our battle simulator. General frequencies are obvious (e.g., movement events occur many more times frequently than other events, more missiles are fired than hits occur). Because of the object–oriented nature of the simulation, we can consider specializations of these classes of events to identify other patterns, for example, all ground vs. air movements, only strike mission movements, and so on.

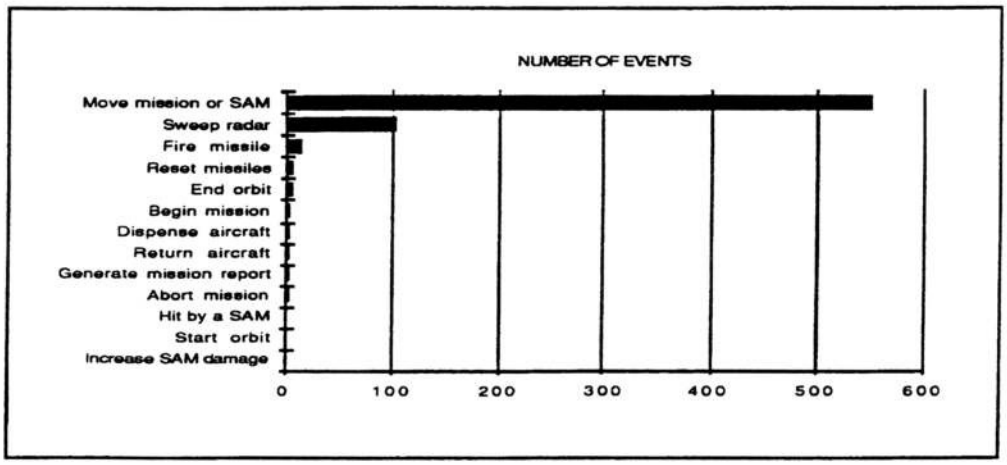

Figure 1. Battle Simulator Event Histogram

It is not only important to determine if events occur with statistically significant frequency. We must also analyze their distribution over time. Many events compressed over a small time frame may indicate important activity, rather than frequently occurring "typical" events. Figures 2 and 3, for example, plot frequencies of several different kinds of events from a typical run of our simulation. While some events occur at a steady state of frequency, others are erratic over time. If we assume frequent or commonplace events are boring, these histograms become a first indicator of what might be interesting or not in a domain. For example in our battle simulator, long–range radar are constantly sweeping, Surface to Air Missile (SAM) sites are always repositioning themselves, and aircraft are always flying point–based ingress/egress routes and so these events, independent of context, are deemed uninteresting and are not reported (see Figure 3). In contrast, missile firings and hits are less frequent but key events. Finally, in domains in which historles can be captured (e.g., previous runs of simulations, previous period financial transactions, etc.), over time we can capture average event occurrences and use these to identify variations from the norm. We also can correlate frequencies of various event classes with one another, detecting potential causal or temporal relations when we are unable to introspect the underlying simulation or events (e.g., noting that aircraft always produce mission reports after they return). Finally, in addition to considering different classes of events, we can analyze events with particular agents, characteristics, and so on, to identify patterns.

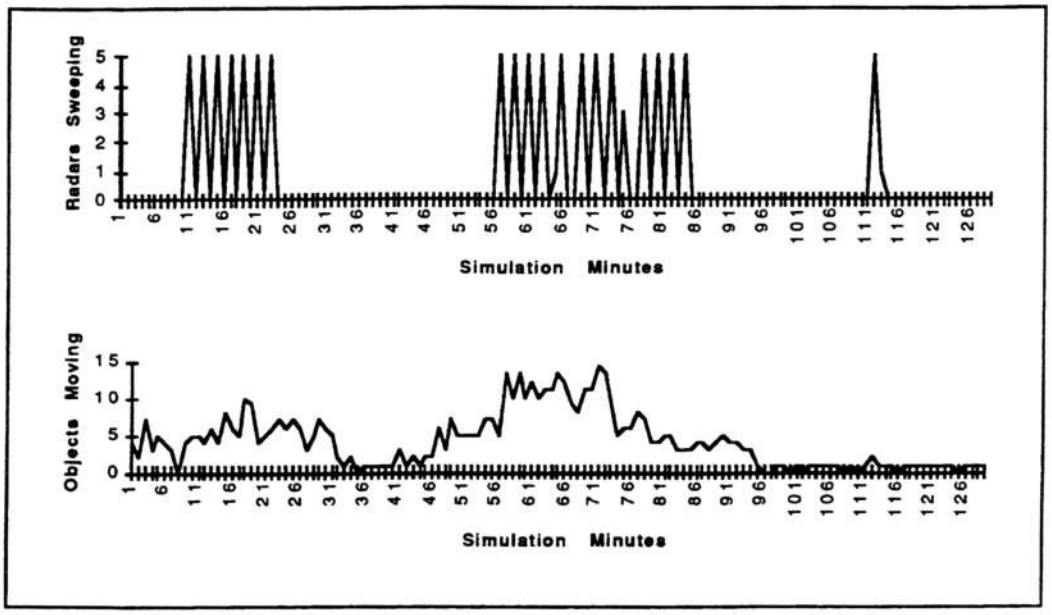

Figure 2. Radar Sweep and Movement Frequency over 129 minutes of simulation time

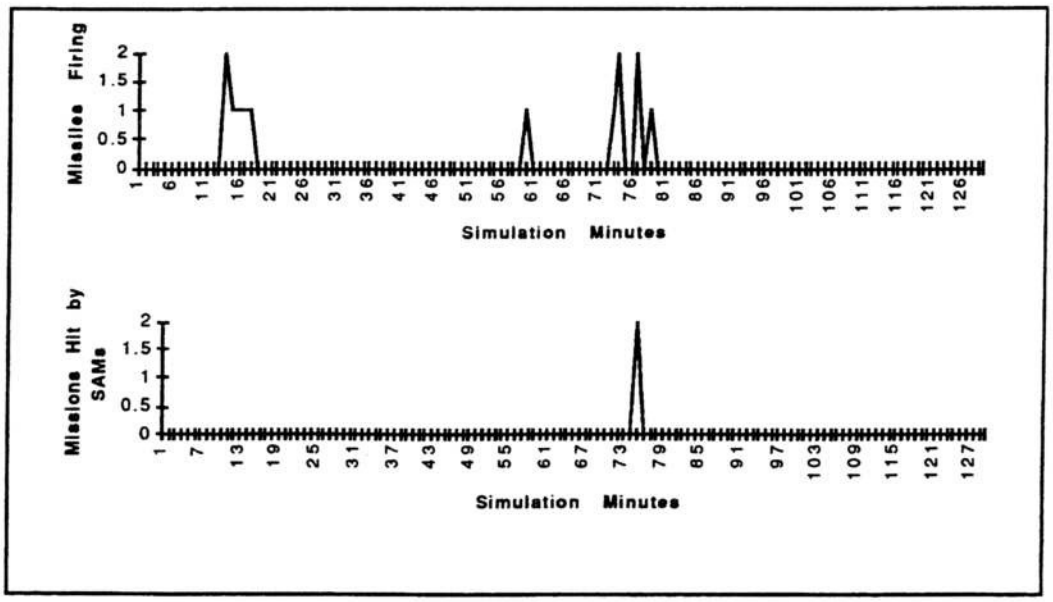

Figure 3. Missile Fire and Hit Frequencies over 129 minutes of simulation time

While the previously mentioned techniques focus on contraining the amount of information to be presented, there are presentational techniques which can help shorten the length of or time required to present information. First, we can exploit context set in previous portions of a dialogue. For example, in our report generator we use of the notions of temporal, spatial, and topical focus (Maybury, 1991) so that after a context is originally set, subsequent references are related to these. For example, we use linguistic constructs such as tense and aspect and temporal adverbs (e.g., "and then", "three minutes later") to convey temporal relations between events without providing exact details of event times. Similarly, spatial adverbials (e.g., "four miles west (from here)") can elide spatial details. Finally, if an event has already occurred and been reported, using the adverbial "again" (as an anaphoric event reference) we can reduce the amount of detail expressed about the event, shortening the resulting presentation.

In contrast to text–only summaries, in multimedia presentations, selecting a particular medium (e.g., text versus graphics) in which to realize information can also result in a savings in the amount of time required to present a given set of content. For example, movement events can be more rapidly displayed and perceived graphically than textually. Thus, while not summarizing the content, this can achieve the effect of reducing the time for (machine) presentation and (user) perception. The use of alternative media to present information (e.g., tabular comparisons versus textual comparisons) requires further investigation, particularly with respect to different summarization purposes and different cognitive tasks.

A very intriguing summarization technique is to utilize the rhetorical structure of the presentation to select what to include or exclude. For example, often an extended description of an object begins with a term definition, in particular, the identification of the superclass and differentia or distinguishing characteristics of the entity. Only then are specifics about the components or different types of the entity detailed. Similarly, in narrative, elaboration of events or background often proceeds events. By identifying the key rhetorical elements (the "nucleus" in RST terms (Mann & Thompson, 1987)), we might begin to identify domain independent rhetorical techniques for summarization. One potential danger is a loss of coherence in the resulting narrative caused by lack of background material or details. More work needs to be done here also.

We have summarized multiple techniques for automated summarization. Each approach has its strengths and weaknesses and our experience is that the selection of techniques is application dependent. Moreover, our experience suggests that a mix of techniques (e.g., performing some statistical analysis of event occurrence coupled with some domain specific knowledge of saliency) performs better than using individual techniques in isolation.

3. A Summary Generator

The above observations result in part from our experiments summarizing events from an object–oriented battle simulation (Anken, 1989) . Our summary generator takes as input time stamped event messages. These several hundred to over seven thousand messages (depending upon how frequently event occurances are sampled) are pruned using a metric based on the frequency, uniqueness, and domain–specific importance of events in the domain. The *agents* of the resulting events are missions (e.g., strike, refueling). These are ranked in order of their frequency of occurrence in the pruned list of events This ranking determines the order of presentation of topics (agent activities) in the final summary. For missions that occur equally frequently, a domain–specific order of importance is used (e.g., offensive air attack > SAM suppression > refueling > transportation; where > indicates greater importance). Having selected the most salient events, the overall organization of the resulting narrative is planned using a communicative–act based text generator (Maybury 1991a, 1992) that sequences the report first by topic (i.e., mission), then chronologically within topic, resulting in a multiparagraph summary report. By encoding user preferences in the constraints of the plan operators, we can also produce (paragraph length) summaries that focus on a particular agents in the simulation (e.g., transportation missions), thus producing a tailored summary for a stereotypical user (e.g., logistician).

4. The Future

While the techniques we have investigated work well in this particular domain, we have not evaluated their performance across other event simulators. Evaluation of summaries is an important concern and should be tied to the purpose of the summarization. One obvious purpose is to convey less information — only the most significant information. This entails trading off detail for conciseness, risking loss of coherence. Doing this correctly also requires reasoning about the inferribility of events and states by the addressee, hence detailed user modeling. Another purpose of a summary may be to convey a given amount of information in less space (i.e., in fewer words or pictures) or in a shorter time period. These latter purposes might be achieved by presentation techniques such as those briefly introduced above.

While we have considered tailoring event selection to particular stereotypical users, several questions remain, such as do different classes of users or characteristics of users require different kinds of summaries, in content or form? Or do different interaction needs (e.g., limited space, time, or attention) require distinct summaries? Another important area for future work is the use of rhetorical and linguistic techniques for summarization (e.g., how temporal, spatial, and other classes of adverbials relate to the eventlstate structures and summarization.) .

In summary, this paper outlines several techniques for summarizing events and aims to stimulate readers to consider experiments combining techniques as we have done. Both analysis of human produced summaries and evaluation of machine generated summaries should help identify the direction for further research to help refine the initial taxonomy of summarization techniques we have outlined here.

5. Acknowledgements

I am indebted to Judy Sider, Sam Bayer, David Day, Angel Acensio, and Nick Pioch for developing and extending the report generator and presentation planning system in which several of the above ideas were explored. Also, I thank Judy Sider who collected event statistics from the battle simulator. Finally, I thank Lynette Hirschman and Judy Sider for comments on earlier drafts.

6. References

Anken, C. S. 1989.

"LACE: Land Air Combat in Eric". Rome Air Development Center TR 89–219.

Goldman, N. M. 1975.

"Conceptual Generation." *Conceptual Information Processing*, editor R. C. Schank. Amsterdam: North–Holland.

Hovy, E. 1988.

"Planning Coherent Multisentential Text." Proceedings of the 26th Meeting of the ACL, Buffalo, NY, June 7–10, 1988. 163–169.

Kittredge, R., A. Polguère and E. Goldberg

25–29 August, 1986. "Synthesizing Weather Forecasts from Formatted Data." Proceedings of COLING–86, The 11th International Conference on Computational Linguistics, University of Bonn, West Germany, 1986. 563–565.

Kukich, K. 1983.

"Design of a Knowledge–Based Report Generator." Proceedings of the 21st Meeting of the ACL, Cambridge, MA, 1983.

Li, P., M. Evens and D. Hier. 1986.

"Generating Medical Case Reports with the Linguistic String Parser." AAAI–86, Proceedings of Fifth Annual Conference on AI, Philadelphia, PA, August 11–15, 1986. 1069–1073.

Lehnert, W. G. 1981.

"Plot Units and Narrative Summarization." *Cognitive Science* 4(1981):293–331.

Mann, W. C. and S. A. Thompson. 1987.

"Rhetorical Structure Theory: Description and Construction of Text Structures." *Natural Language Generation*, editor G. Kempen. 85–95. Dordrecht Martinus Nijhoff.

Maybury, M. T. 1991a.

Planning Multisentential English Text using Communicative Acts. Ph.D. dissertation, University of Cambridge, England, June, 1991. Available as Rome Air Development Center TR 90–411, In–House Report, December 1990 and as a Cambridge University Computer Laboratory TR–239, December, 1991.

Maybury, M. T. 1991b.

"Topical Temporal and Spatial Constraints on Linguistic Realization" *Computational Intelligence: Special Issue on Natural Language Generation*. Volume 7(4). 266 275.

Maybury, M. T. August, 1992.

"Communicative Acts for Explanation Generation" *International Journal of Man–Machine Studies*. 37(2), 135–172.

McKeown, K. 1985.

Text Generation. Cambridge University Press.

Meehan, J. R. 1976.

The Metanovel: Writing Stories by Computer. Ph D dissertation, Yale University TR 74, New Haven, CT.

Paris, C. L. 1988.

"Tailoring Object Descriptions to a User's Level of Expertise." *Computational Linguistics: Special Issue on User Modeling* 14(3):64–78.

Related Disciplines – discourse analysis and use I

Rapporteur: Ines Busch–Lauer

Jerry Hobbs:

Summaries from Structure

Illustrating the notion of adjacency of pairs of words, compound nominals, syntax and compositional semantics, Jerry R. Hobbs demonstrated that interpreting texts includes explaining the adjacency of clauses, sentences and larger segments of discourse.

Based on the fact that the source of discourse structure is precisely the same as the source of intrasentential syntactic structure (a predicate–argument relation); and rather of a compound nominal in case of longer stretches of text, an abductive framework was developed for syntax, compositional semantics and local pragmatics, wherein to parse and interpret a sentence W is to prove the expression `( - e) S (W,e)` meaning that there is a situation or eventuality e such that the string of words W is a grammatical sentence describing or conveying e (the assertion of the sentence). Both hypotactic and paratactic coherence relations are captured in this framework.

According to the proposed model the text "Bank Robbers" was analysed, resulting in a tree–like structure of the text with five types of coherence relations:
Explanation, Ground–Figure, Parallelism, Contrast, Occasion.

The report provoked a lively discussion focussing on:

1. Procedures of getting a summary out of this proposed tree–like model from a bottom–up or top–down approach; Procedures of getting summaries of linear and dynamic texts.
2. The coherence relations in the sample text (sentences 1 and 2 in relation to 3 to 9; 4a to 4b–9); the role of propositions for summarizing.
3. Computational realisation of the model, e.g. for large–scale, complex text structures.
4. The relevance of the knowledge base for summarizing. What role does the world knowledge of the summarizer play in the process of summarizing? Is the knowledge of the user of summaries to be considered in this process? There is the need to adapt summaries to users' knowledge and purpose.

Summaries from Structure

Jerry R. Hobbs
Artificial Intelligence Center
SRI International

Structure from Adjacency

To understand our environment we seek the best explanation for the observable facts we find there. Similarly, to interpret texts we seek the best explanation for the "observable" facts that are presented in the text. This view can be cashed out computationally by taking the interpretation of a text to be the most economic abductive proof of the logical form of the sentences in the text, where "abductive" means that assumptions are allowed at various costs (Hobbs et al., 1993).

Among the observable features of our environment that we seek to explain is the adjacency or proximity of objects; this generally escapes our notice except when it is out of the ordinary, as when we enter a room and see a chair on top of a table. A similar situation obtains in language. A text is a string of words and one of the features of the text that requires explanation is the adjacency of pairs of words or larger segments of text.

The simplest example of this is provided by compound nominals. When we see the phrase "turpentine jar" in a text, the interpretation problem we face is finding the most reasonable relationship in the context between turpentine and jars, using what we know about turpentine and jars. In many cases, the relationship is one arising out of one of the nouns itself, as in "oil sample", where the relation between oil and the sample is precisely the "sample of" relation.

Syntax and compositional semantics can be seen as arising out of the same need to explain adjacency. When we see the pair of words "men work", we need to find some relation between them. The hypothesis that sentences have syntactic structure amounts to the acceptance of a set of constraints concerning the relation that can obtain between two words or larger stretches of text. In this case, the constraint is that the second word itself provides the relationship. The men have to be the agent of the working. Whereas in the case of "oil sample" "sample" provides a possible relationship, in the case of "men work" "work" provides the obligatory relationship. (This is not quite true; metonymy is possible, so that the second word need only provide a relationship between the event and something functionally related to the the first word.) 1

The tree structure of sentences arises from the fact that the adjacency relation can be between larger segments of text than simply single words, where the segments have their own internal tree

structure resulting from adjacencies. For example, in

John believes men work.

we don't seek to explain the adjacency between "believes" and "men". Rather we first explain the adjacency between "men" and "work", and only then the adjacency between "believes" and "men work" (or the adjacency among "John", "believes", and "men work", depending on your view of the structure of the clause.) This grouping occurs even in the absence of syntactic constraints. Consider the two compound nominals, "Stanford Research Institute" and "Cancer Research Institute". In the latter, we must first find the relationship between cancer and research, and then find the relationship between cancer research and the institute, whereas in the former, we group "research" with "institute" and then "Stanford" with "research institute".

In order to explain adjacencies between segments of text larger than one word, we need to have an idea of the principal information conveyed by the segments. For example, in "research institute" the reference to the institute is, in a sense, more primary than the reference to research. A research institute is a kind of institute, rather than a kind of research. Similarly, in "men work" the information about some entities working is more primary than the information about the entities being men. Therefore, as we compose larger and larger segments of text, we must have some sense of the primary information conveyed by the composite segments.

The rules of syntax and compositional semantics are a set of constraints on how segments of text can be grouped together and on what the primary information conveyed by the composite segment is. At the level of the main clause, the primary information is often what is conveyed by the main verb and/or by adverbials, although this can be overridden by such factors as intonation, newness, topicalization, and so on. We may call this primary information, however it is determined, the *assertion* of the clause. It is, in a sense, a summary of what the sentence conveys.

The source of discourse structure is precisely the same as the source of intrasentential syntactic structure. Two phrases, or clauses, or sentences, or larger stretches of discourse are adjacent in the discourse, and this fact requires explanation. A relation between them must be found to explain that adjacency. Whereas in the syntactic structure of sentences the relation among adjacent elements is most commonly a predicate–argument relation, the case of larger stretches of unrestricted text is more like the case of compound nominals, in that the relation that explains the adjacency can in principle be any plausible relation between the situations described by the segments of text.

As with the case of syntax, while we compose larger and larger segments of text, we must have some sense of the primary information conveyed by the composite segment. We must be able to

specify the ''assertion'', or equivalently, a summary, of a supraclausal segment of text. This is both harder and easier than in the case of compound nominals. It is harder in that whereas it is always the second noun of a noun–noun pair that is primary, with two supraclausal segments it may be the first or the second, or the primary information may arise equally out of both.

It is easier in that a smaller number of relations can obtain between two situations or eventualities described in supraclausal segments than can obtain between two nouns. For compound nominals, Downing (1977) and others have convincingly argued that the relation between the two nouns can be virtually anything, given the right context. On the other hand, Levi (1978) argues convincingly that the vast majority of the relations can be viewed as instances of no more than a dozen or so different abstract relations, such as predicate–argument, function, containment, and so on.

Similarly, it is possible that the relation between two adjacent supraclausal segments of text can be anything at all in the right context. The hearer must simply figure out the most plausible relation between the situations described and the most plausible assertion or summary of the composite segment. Unfortunately, such a view of discourse coherence gives no guidance as to what the assertion or summary of the composite segment is.

But overwhelmingly, the relation between supraclausal segments can be viewed as an instance of one of three broadly construed abstract relations–causality, the figure–ground relation, and similarity. I will refer to these as *coherence relations*. A theory of discourse coherence and discourse structure that recognizes this fact about discourse must develop characterizations of each of these relations, explicate the various classes of instances of each relation, and for each of these classes, define the assertion or summary of the composite segment. This is what I have tried to do in previous work (e.g., (Hobbs, 1985) and what I will try to recast into an abductive framework in this paper.

While I will focus on the way these three relations relate supraclausal segments of text, they can also relate material within single clauses. For example, elements of a list exhibit the similarity required of parallelism (cf. Polanyi, 1988) . In the sentence,

A car hit a jogger in Palo Alto last night.

there is an implicit causal relation between the jogging and the hitting. These relations go beyond what is given to us by compositional semantics.

While I have focused, and will continue to, on the *interpretation* problem, it is important to keep in mind that interpretation and generation are intricately interrelated. A speaker seeks to generate utterances that will be understood. When two segments of coherent discourse are uttered in

sequence, it is because the speaker expects the hearer to recover the relation that is intended to be conveyed by the adjacency. Conversely, the hearer must often reason about what the speaker is trying to achieve and the other ways in which the speaker might have chosen to achieve it, in order to determine the best interpretation for a stretch of discourse.

From Coherence Relations to Structure and Summaries

In Hobbs et al. (1993) a unified abductive framework for syntax, compositional semantics, and local pragmatics is presented. In this framework, to parse and interpret a sentence W is to prove the expression

$$(\exists\, e)S(W, e)$$

meaning that there is a situation or eventuality e such that the string of words W is a grammatical sentence describing or conveying e. e is the assertion of the sentence. We will take this as our starting point for a treatment of discourse structure.

The tree–like structure of discourse can be captured with two axioms:

$$(\forall\, w, e)S(w, e) \supset Segment(w, e)$$

$$(\forall\, w_1, e_1, w_2, e_2, e)Segment(w_1, e_1) \wedge Segment(w_2, e_2) \wedge$$

$$CoherenceRel(e_1, e_2, e) \supset Segment(w_1 w_2, e)$$

The first axiom says that a sentence w describing an eventuality e is a coherent discourse segment describing e. The second says that if two segments w_1 and w_2 describe the eventualities e_1 and e_2, respectively, and e_1 and e_2 are related by some coherence relation, then the concatenation $w_1 w_2$ is a coherent discourse segment.

The variable e in the second axiom is the assertion or summary of the composite segment $w_1 w_2$. It is determined by the assertions of the constituent segments, e_1 and e_2, together with the relation that holds between them by virtue of which $w_1 w_2$ is itself a coherent discourse segment.

To prove $CoherenceRel(e_1, e_2, e)$ is to explain the adjacency of w_1 and w_2.

To interpret a text W is then to prove the expression

$$(\exists\, e)Segment(W, e)$$

meaning that W is a coherent segment of discourse conveying or describing the situation or eventuality e. e is the assertion or summary of W. Thus, one of the products of an interpretation of this sort is a summary of the text derived from its structure.

To explicate a theory of discourse coherence and discourse structure along these lines is to specify the various ways in which

$$CoherenceRel(e_1, e_2, e)$$

can be established, including what e is.

The common distinction between hypotactic coherence relations, with dominant and subordinate component segments, and paratactic coherence relations is easily captured in this framework. If the relation is hypotactic, then e is either e_1 or e_2, corresponding to whether w_1 or w_2 is dominant. If the coherence relation is paratactic, then e must be computed from e_1 and e_2.

In the talk, I will examine five examples of coherence relations and their corresponding summaries.

1. **Explanation:** This is defined in terms of causality. The relation is hypotactic. The explanandum is dominant and thus contributes the assertion or summary; the explanans is subordinate.

2. **Ground–Figure:** The definition of this relation requires us to have axiomatized the notion of a *system*, that is, a set of entities and relations among them, and the notion of an entity external to the system being, in some perhaps metaphorical sense, *at* an element in the system. The relation is hypotactic. The figure is dominant and contributes the assertion or summary; the ground is subordinate.

3. **Parallelism:** The definition of this relation involves recognizing the similarity of entities. The text segments must assert the same properties of similar entities. The relation is paratactic. The summary of the composite segment is the generalization of which the assertion of each segment is an instance. Elaboration is a limiting case of this coherence relation, where the entities are not merely similar but identical.

4. **Contrast:** This relation is defined much like parallelism. The text segments must assert contrary properties of similar entities. The relation is generally hypotactic. Usually, the second segment is the dominant one, but this can be overridden.

5. **Occasion:** This is a weak form of causality or a strong form of temporal succession. It

involves the first text segment describing an event or situation that sets up the occasion for the event or situation described by the second segment. The relation is paratactic. The summary or assertion of the composite segment is a coarser–grained description of the sequence of state changes described by the component segments.

A summary of the entire text can be derived in this fashion. However, it should be kept in mind that it may not be the ''summary'' that is required for any particular application.

Footnote

1] This view of syntax emerged from a conversation with Mark Johnson.

References

Downing, Pamela, 1977. ''On the Creation and Use of English Compound Nouns'', *Language*, vol. 53, no. 4, pp. 810–842.

Hobbs, Jerry R., 1985. ''On the Coherence and Structure of Discourse'', Report No. CSLI–85–37, Center for the Study of Language and Information, Stanford University.

Hobbs, Jerry R., Mark Stickel, Douglas Appelt, and Paul Martin, 1993. ''Interpretation as Abduction'', *Artificial Intelligence*, No. 63, pp. 69–142.

Levi, Judith, 1978. *The Syntax and Semantics of Complex Nominals*, Academic Press, New York.

Polanyi, Livia, 1988. ''A Formal Model of Discourse Structure'', *Journal of Pragmatics*, Vol. 12, pp. 601–638.

Related Disciplines – discourse analysis and use II

Rapporteur: Woojin Paik

Annely Rothkegel:

Abstracting in the perspective of producing a text

Annely Rothkegel (University of the Saarland) reported on theoretical principles of a text production model. The presenter noted that abstracting or summarizing is a specific sub–type of text production. The presenter also said that strategies of producing the mini–text controls the processing of the maxi–text. Representing processes of abstracting is described as representing text production processes.

The presenter showed that there are three levels of text space. They are:

a) text content (information);
b) text function (presentation of information); and
c) text form (sequencing of (b(a)).

She also explained that three characteristics of text actions (TAs): change of states (CHANGE); propositional content (PROP); and explicit markers (LANG) relate to form a predicate–argument structure.

In conclusion, the presenter remarked that summarizing is a sub–type of general writer's activities which can be described in terms of TAs. TAs are an instrument for representing human strategies of the text composition. In addition, the explicit formulation of the relationship between function, content, and linearization allows the construction of a computer model which may be applied for investigating and/or supporting human text production.

In discussion, there was an inquiry about the relative position of the presenter's theory in terms of Searle's notion of perlocution. In response to a question about the difference between German and English, the presenter responded that German chains of connectors are more explicit but the conceptual level is the same. There was a question regarding the contraint of the source text structure on the summary text structure, and the options of transforming the source text structure to the summary text structure. Regarding the comment about the general schema for texts, the presenter responded that the schema is very flexible in her model. One audience member commented that the presenters's model is about text reception not text production since the result

is based on her own introspection. The response was that it is reconstruction on the basis of what regularities can be found on which to reconstruct.

Livia Polanyi:

Linguistic Dimensions of Text Summarization

Livia Polanyi (Linguistics, Rice University) reported on the complexity of human discourse and the Linguistic Discourse Model (LDM). The presenter asserted that the discourse is structured and should properly be considered part of core grammar: "it is not soup", she claimed. Subsequently, using a sample naturally occuring multi–interaction, the presenter showed that two utterances which are sequentially ordered in texts may not express information relevant to the same interactional, social, genre or modal context.

The presenter explained that LDM has two basic elements of discourse formation. One is the discourse constituent (dcu) which is propositional and the other is discourse operator which is non propositional. Dcus are equivalent neither to clauses nor to sentences i.e. sentences are not surface reflex of dcus, but are minimal "semiotic reflexes" of a minimal proposition in a modal context. Discourse is constructed through recursively embedding and sequencing dcus to one another. Types of discourse structure which are formed by recursively coordinating and/or sequencing elementary dcus are: lists; subordination structures; binary structures; genre units; and speech events. The presenter proposed that an interpretative component which consists of a set of model theoretic dynamic discourse unit representations and a formal mechanism to describe accessibility relations among these representations is necessary for discourse processing machine.

Turning to a discussion of salience in discourse, the presenter discussed some work on storytelling which maintains that a storyteller reveals relative importance of propositions linguistically, by phonetically, lexically, syntactically, and rhetorically marking information considered salient by the teller in a manner which allowed that information to be "noticeable" to the recipient. The general rule, she claimed, is that the more distinctive the encoding form a proposition receives, the more weight a listener knows to attach to it. In composing a summary, it is possible to take the most highly marked information and assemble it into a summary using basically surface levels linguistic clues and information returned by the LDM discourse parse mechanism.

The presenter then described 'Summa Auto Summarization System' which is being developed by AND Software in the Netherlands and which is based on some of her earlier work. The system uses lexical clues; clue verbs; and syntactic information together with available text processing tools such as word frequency lists, thesauri, dictionaries etc. The system generates 1 sentence summaries of 4 paragraph newspaper articles of which 65% are acceptable, for Dutch texts. She

pointed out that the tools available for Dutch are not as well developed as those for English and that the rate of acceptability would be higher for English texts.

In the discussion, there was a comment that an example based on the presenter's theory seems to be organized chronologically rather than causally. This was answered by the presenter who maintained that though causal relations are often associated with chronologically arranged event structures in narrative, such causality is a second order effect. The primary organizing principle of narrative is chronological ordering of event information — where events are punctual propositions encoded in syntactically distinct ways.

Abstracting from the perspective of text–production

Annely Rothkegel

University of Saarbrücken
rothkegel@coli.uni–sb.de

Dagstuhl, December 1993

1. Introduction

This abstract introduces some theoretical principles of a text production model (Rothkegel 1993) which will be applied to an authentic text in my talk. The approach is oriented towards both text pragmatics and computational linguistics. As a starting point I will rely on the relationship between the "abstract" and the "referred text". Specifically, we could distinguish three kinds of relationship between the two:

(a) The referred text is pre–existent but not co–present with the abstract.
(b) The referred text and the abstract are co–present.
(c) The abstract is pre–existent whereas the referred text has not been produced yet.

The question arises as to whether there are some aspects connecting the three cases. Common to all the three cases is the fact that a central idea which is related to another text is designed, developed and contextually integrated in a very concentrated manner. Generally, this corresponds to the development of the text theme and the text function in every text. In case (a) the information about the text theme is derived from the referred text whereas functional aspects (such as reference to the readers) come from the knowledge of the writer of the abstract. Case (c) represents a kind of plan of what will be realized in a text in future. Case (b) covers (a) or (b) or a mixture of both.

The connecting principle is that of construction. New information about an object (here the object is a text) is constructed on the basis of a specific text structure. Principles of construction play a central role in the area of text production. We can therefore enquire whether some phenomena of abstracting may be investigated in terms of general conditions of text production.

2. The model of text production

In the following I will sketch a text production model which focuses on two main aspects: the

construction of the text structure and the representation of structure–building activities of writers (or of a computer system).

(i) Construction of the text structure

The building up of the text structure refers to the text as a whole entity. Text structure is considered to be a phenomenon per se which is distinct from knowledge structure on the one hand and from sentence structure on the other. It is assumed that there are some extra–linguistic effects of the text structure which consist in making available an instrument for constructing a new informational coherence. Two levels should be considered:

(1)

The level of content and function which provide the choice of information units. This choice is manifest in terms of both the selection and development of themes. Principles of theme development are dependent on the formal properties of a text.

(2)

The level of text form which provides different kinds of composition. This level takes into account that a text is a medium of language which is subject to conditions of linearity. The information units are to be realized in terms of a sequence. This means that they must each be connected to another in a specified manner and in a particular order. Linguistically, principles of composition are reflected in phenomena of connectivity within the text.

(ii) Representation of the writer's activities

The building up of the text structure is understood as a process which is performed by writers. The interesting question here concerns problems with respect to how these activities are related to the text levels mentioned above. Therefore in this section (ii), I will consider some aspects of section (i) from a dynamic perspective. Seen from this perspective, the selection, development and sequential connection of the thematic units are all objects of human cognitive processing. This dynamic view of information corresponds to a pragmatic view which was been argued for by Winograd &Flores (1986) with regard to "social cognition" and the design of language oriented computer systems.

A theoretical basis for this approach is provided by the speech–act theory (Searle 1969), especially by process–oriented modification (Sbisa 1987). In this view, linguistic actions are described with respect to the change of states which are effected by them. This concept is very helpful for modelling text production insofar as the construction of a new coherence of information is understood as the change of an informational state (of the writers as well as of the readers!).

A special point to be noted is that linguistic actions (in the sense of speech–act theory) are

determined as social behaviour which is bound to commitments of the speakers/writers. Within the framework of text production, the correspondent idea concerns the fact that information in a text always exists as information which is embedded in a communicative situation. Thus it may be a part of a description, a report, a narrative or an argument. The different situations commit the speaker/writer to different consequences. In a report, for instance, a writer commits himself/herself to the truth and order of the events that are reported. If the writer chooses an different order (different text structure) he/she has to introduce specific linguistic markers. Or, the description of a fact entails some other consequences than the application of a fact as an argument for the pro– or contra–position of the writer. Therefore it seems quite reasonable to interpret the building up of a text structure in terms of linguistic actions. We can take advantage of two ideas. The writer's activities

- can be specified on the basis of text structures (with respect to the envisaged goal of the text),
- can be represented in a formal way on the basis of speech–act theory (this allows some advances between a model of human text production and computer simulation of writing).

In the following, text producing actions will be designed as text actions (TA). The speech–act theoretical distinction between illocution, propositional content and locution is then reflected in correspondencies which focus on the aspect of change. In this view TAs are generally characterized as changes of states (CHANGE). These changes provide some new informational states which are represented in terms of the propositional content (PROP) and which are accompanied by explicit markers which indicate the kind of change or by implicit relationships which are to be infered (LANG).

The three components are related in such a way that they themselves form a predicate–argument structure (cf. Hafele 1979, Allen 1987):

```
[1]      TA: CHANGE ( PROP, LANG)
```

[1] is the basic formula for further specifications:

Concerning CHANGE:

Changes of the informational state refer to the three central aspects of text structure: text function (FUNC), text theme (THEM) und text sequence (SEQU).

```
[2]      CHANGE: < FUNC, THEM, SEQU >
```

CHANGE (FUNC) organizes the integration of information into specified text functions. Some of

the corresponding text actions are DESCRIBING, REPORTING, NARRATING, ARGUING, etc.
CHANGE (THEM) organizes information according to theme–processing schemata. Examples of corresponding text actions are: THEMATIZING, GENERALIZING, DETAILING, etc (cf. Rothkegel 1984).
CHANGE (SEQU) organizes information by sequence. Here we have two variations: global and local sequencing (S–GLOB, S–LOC).

Global sequencing organizes information with respect to the completion of global text schemata. In the case of functional aspects there are e.g. complete lines of argumentation or narrative patterns (PATT). With respect to the text theme there are schemata concerning the relationsship between the main theme and some side themes or between the main theme and some part–themes (cf. Brinker 1990). The general text action for realizing this structure is CHAINING: a list of propositions is connected in a particular sequence.

Local sequencing organizes information with respect to neighbouring units. The general text action is PAIRING. PAIRING is realized according to two different strategies: repetition/maintenance (MAINT) and shift to another category (SHIFT). Both MAINT and SHIFT refer to the functional level as well as to the thematic level. Examples for MAINT are CONTINUING, REPEATING, etc. Examples for SHIFT are CONTRASTING, NEGATING, etc.

```
[3]     CHANGE (SEQU):
        S-GLOB/CHAINING (PATT (FUNC, THEM)),
        S-LOC/PAIRING (MAINT (FUNC, THEM), SHIFT (FUNC, THEM ))
```

Concerning PROP:

The information concerning the content is represented in terms of propositions (predicate–argument structures). The structural slots can be filled by lexemes or concepts as well as by general semantic roles (cf. de Beaugrande & Dressler 1981) or by text roles which are domain specific (Rothkegel 1993). The component of PROP is responsible for the link between the sources of information and the linguistic units. A possibility of organizing the access to these sources may be the concept of frames (Minsky 1977, Tonfoni 1990). It provides an organizational structure (FRAM) for very different knowledge sources which are relevant for the content of a text. As far as the process of abstracting is concerned, the refered text or the future text play a dominant role. Nevertheless the categories of the slots should be determined by the expectations of what is interesting in an abstract. Lastly, these slot categories represent the argument roles in the propositional structure:

```
[4]     FRAM-i: { ROL-i1, ..., ROL-in }
        FRAM-k: { ROL-k1, ..., ROL-kn }
```

```
etc.
->
PROP-a: PRED ( ROL-i, ... )
PROP-b: PRED ( ROL-k, ...)
```

Concerning LANG:

Linguistic properties of text connectivity are combined with functional and/or thematic units. This linkage can or must be realized in an explicit form which is represented in LANG. It is assumed that there are some inventories of lexicalization which are dependent on text patterns or which are preferred in specified domains.

3. Summary

The text production model

- provides an instrument for the research of some text–oriented strategies of writers; within this framework, abstracting is a sub–type of general writer's activities which can be described in terms of text actions;
- is a basis for the representation of specified instructions for text production; the explicit formulation of the relationship between function, content and linearization allows the construction of a computer model which may be applied for investigating and/or supporting human text production.

References

Allen, J.F., 1987.
Natural Language Understanding. The Benjamins/Cummings Publ. Comp., Menlo Park.

de Beaugrande, R.–A. & Dressler, W., 1981.
Einführung in die Textlinguistik. Niemeyer, Tübingen.

Brinker, K. 1990.
Linguistische Textanalyse. Eine Einführung in Grundbegriffe und Methoden. E–Schmidt–Verlag, Berlin.

Hafele, J., 1979.
Der Aufbau der Sprachkompetenz. Untersuchungen zur Grammatik des sprachlichen Handelns. Niemeyer, Tübingen.

Minsky, M., 1977.
Frame System Theory. In: Johnson–Laird, P.N. & Wason, P.C. (eds), *Thinking: Readings*

in Cognitive Science, 1977: 355–377, Cambridge University Press, Cambridge.

Rothkegel, A., 1984.

Thematisieren als Texthandlung. In: Krenn, H. & Niemeyer, J. & Eberhardt, U. (eds), *Sprache und Gesellschaft*, 63–73. Tübingen, Niemeyer.

Rothkegel, A., 1993.

Text Knowledge and Object Knowledge. Pinter, London.

Sbisa, M., 1987.

Speech acts and context change. In: Ballmer, Th. & Wildgen, W. (eds), *Process Linguistics. Exploring the processual aspects of language and language use, and the methods of their description*, 1987: 252–280. Niemeyer, Tübingen.

Searle, J., 1969.

Speech acts. Cambridge University Press, Cambridge.

Tonfoni, G., 1990

Text representation systems. Eurographica, Helsinki.

Winograd, T. & Flores, F., 1986.

Understanding computers and cognition. A new foundation for design. Addison–Wesley Publishing Company, Reading.

Linguistic Dimensions of Text Summarization

Livia Polanyi

Rice University, Houston TX
and Institute for Research on Learning, Palo Alto, CA

December 9, 1993

Text summarization: a complex problem

Coming up with a general solution to the problem of text summarization was identified in the early 1970's as one of the core tasks of computational linguistics and AI. Twenty years later relatively little progress has been made on developing robust, domain independent approaches to extracting the key ideas from a text and assembling them into a compact, coherent account of the source. Summarization remains an extremely difficult and apparently intractable problem. In my view, this is because discourse which takes information from other modules of the grammar as input, and returns as output information needed to build representations of the meaning of the discourse, emerges from highly complex and ill–understood interactions among sentential prosodic, syntactic and semantic knowledge, knowledge of discourse structuring conventions, and world knowledge. Despite the primitive state of our understanding of discourse, I believe that there is a great deal to be gained for summarization from understanding the linguistic structure of the texts to be summarized. In this brief essay I would like to discuss from a linguistic perspective, the properties of the discourses which human beings regularly produce and decode.

2 Discourse Context

Texts are complex semantic objects. Much as we might like to consider the texts which we might care to summarize as "cohesive", "coherent", "monologic" and structures, it is unfortunately true that an important property of naturally occurring discourse – even military casualty reports or simple weather reports – is that successive semiotic gestures (1) need not convey information or reaction relative to the same abstract semantic model. For example, a page of a newspaper may tell a number of independent "stories" (or parts of stories) each modelling a separate object world. Even within "one" overall story, a number of separate "worlds" may be invoked: one telling about what generically occurs in a given situation, another telling a story about some event which happened in the past which may be relevant to the "news" being told, while still others may tell that "news" – some event of a previous day, perhaps – from the viewpoints of a number of individual characters. Therefore, in processing successive linguistic utterances, we must build up

representations of each abstract semantict/pragmatic world while computing the relationships which obtain among those worlds.

Our experience in discourse modelling suggest that at least the following contexts are crucially necessary in utterance interpretation:

- *Interactions* which situate Real Speakers in Real locations in a Real time in the "real world" (2). (3)

- *Speech Events* which situate participants in roles in a socially constructed world of activities, tasks and actions with their attendant definition of objects, times and space and the rights, intentions and obligations of persons.

- *Modal Contexts* which represent the attitude (real or assumed) of the speaker (real or assumed) towards the content of the utterance as reflected in clausal mood, polarity, empathy, and point of view.

- *Genres* which set up discourse worlds with times, locations, objects and concerns local to the linguistic entity.

When one examines a stretch of discourse with an eye to which contexts each utterance is relevant, it turns out that while there is no guarantee that propositions with similar utterances will cluster in typical formations, in practice, such clustering is the norm. So, for example, while newspapers stories may be broken up with the beginning of a story on one page and its continuation on a subsequent page, the propositions making up the story will be arranged in a conventional way: information giving a general overview of the story will come first (4), and then a narrative line will be developed giving events in order perhaps through the device of giving information in reported speech. Which information from which context must figure in the summary? The entire mainline narrative? Some of the events of the mainline? (Parts of) the mainline narrative plus "important" aspects of the flashed segment? Remarks on the relevance of the events as reported in quotations or free indirect speech? How would one decide? Are there general procedures for making such a decision or must they be made on an ad hoc, case by case basis?

Similarly, interactions are structured in a characteristic manner (5) as are the Speech Events which they contextualize.

3 Discourse Units

Discourse does not usually consist of an unstructured collection of unrelated utterances. Rather,

discourse is characterized by arrangements of information patterened into conventional structures,which means that we may characterize discourse in terms of discourse unitswhich can be described by rules of well–formedness.

In our work in discourse modelling we have identified two basic units of discourse formation: the basic discourse construction units are the *discourse operator* (6) and the *discourse constituent unit (dcu)*. Discourse is constructed through recursively embedding and sequencing dcus to one another. Elementary dcus are contextually indexed representations of information encoded in minimal proposition carrying linguistic structures. (7). Elementary dcus participate in and constitute more complex types of dcus described by their own rules of syntactic and semantic or pragmatic wellformedness. These units include:

- *Sentences* formed according to the rules of sentential syntax
- *Lists* such as narratives and other ordered and unordered sequences
- *Elaborations* including explanations, multiclausal descriptions of mentioned items, direct discourse and similar semantically related phenomena
- *Logical structures* such as IF/THEN, x BUT y units
- *Rhetorical structures* such as thesis/support structures (8)
- *Binary structures* such as Interruptions, Repairs and Adjacency Pairs (greetings, apologies/acceptances)
- *Genre units* such as stories and arguments which include expected information in an expected order
- *Speech Events* such as casual chat, funerals.

4 Discourse processing

Sentential syntax, semantics and prosody guide the segmentation of text into "naked" dcus. World knowledge and linguistic knowledge about discourse activities and their structures as well as specific knowledge about the state of the discourse enables language users to assign appropriate context indices to these "naked" propositional structures to form elementary indexed dcus. Relation among these dcus as they are deployed sequentially in language discourse construction can be modelled at a purely linguistic level – not as a psychological or cognitive process – human

discourse processing can be modelled using Context Free rules which build dcus from dcus and captures discourse hierarchical structure modelled as a stack by Reichman and Grosz and Sidner, as an Open Right Tree by Polanyi and Scha, and as a Tree with an Open Right Edge (the "Right Frontier") by Webber. All of these frameworks share the assumption that discourse hierarchical structure figures prominently in the interpretation of pronomial, locative and temporal anaphora and it is not unreasonable to expect that these researchers would agree that any scheme to automatically summarize discourse must have access to segmentation and parsing devices which would allow for the representation of hierarchical structure.

5 Discourse Interpretation

In order to provide a formal mechanism for incremental interpretation of a series of propositions belonging to a given (set of) contexts, we propose that each discourse processing machine must be equipped with an interpretative component to consist of a set of model theoretic *dynamic discourse unit representations* enriched with inference capabilities and a formal mechanism to describe accessibility relations obtaining among them. This machine would output a set of representations corresponding to the "meaning of the text" in the sense of the states of affairs asserted in relation to each Interactional, Speech Act, Genre and modal context of the source text. These representations, in turn, function as input for the summarization process – since only a portion of the information in each would be included in the target summary. Exactly which information should be included and which left aside to be determined by weighted iunteractions computed among the intentions of the speaker (as shown by the rhetorical marking of the source as discussed in Polanyi 1989), the interests of the recipient (deriving from a model of the user), and the structuring of the text.

6 Text summarization: Beyond the State of the Art?

A formal linguistic theory such as the one we have sketched is AI–Incomplete. An automatic summarization engine which could accept an input text and output a summary of the source text would need access to the information necessary to construct an adequate semantic representation. This would necssarily require a complex model of world knowledge. It is my considered opinion that modelling the world knowledge component of human cognitive functioning in anything other than a toy domain lies far beyond our current capabilities.

7 Footnotes

(1)

Freestanding words, exclamations, phrases or clauses in written language augmented by non–verbal gestures in interactive talk.

(2)

Possibly a fictive or assumed world, cf. literature

(3)

The Interaction Context corresponds to Kaplan contexts needed to interpret indexicals such as "I", "now", "here". (Kaplan [6]).

(4)

What, Who, Why and When

(5)

SIGNON – (DO SPEECH EVENT) – SIGN OFF

(6)

Operators are non propositional in nature. They give information about the state of the discourse and may give supplemental (i.e. affective) clues about propositional information. There are several classes of *operators* including *discourse PUSH and POP markers (cue words, vocatives, greeters*, etc.

(7)

The dcu is a semantic unit corresponding to a minimal predication interpretable in a modal context. The surface reflex of an elementary dcu is most typically a deep clause

(8)

Treated most explicitly by Hobbs and Mann and Thompson within CL and by Longacre in descriptive linguistics

8 References

1. Grosz, B. and C. Sidner. 1986.
Attention, Intention and the Structure of Discourse. *Computational Linguistics*: 175–204.

2. Hinrichs E. and L. Polanyi. 1986.
A Unified Account of a Referential Gesture in *The Proceedings of the Parasession on Pragmatics*. Chicago: Chicago Linguistics Society.

3. Hirschberg, J. and J. Pierrehumbert 1986.
The intonational structuring of discourse. *ACL* 24: 136–144

4. Hobbs, J. 1985.
On the Coherence and Structure of Discourse. *Center for the Study of Language and Information Technical Report* 37, Stanford, CA

5. Johnson–Laird, P. 1983.
Mental Models. Harvard University Press.

6. Kaplan, David. 1989.
in J. Almog,J. Perry, and H. Wettstein (eds.). Themes from Kaplan. NY: Oxford University Press.

7. Litman, D. 1985.
Plan Recognition and Analysis. *TR 170*. Rochester NY: Dept. of Computer Science, University of Rochester.

8. Mann, W. C. and S. A. Thompson. 1987.
Relational Propositions in Discourse. *Discourse Processes* 9 (1): 57–90.

9. Polanyi, L. 1986.
The Linguistic Discourse Model. *BBN Technical Report 6409*. Cambridge MA: Bolt, Beranek and Neman Inc.

10. Polanyi 1987.
Keeping it all Straight: Interpreting Narrative Time in Real Disourse. *WCCFL* 6: 229–245.

11. Polanyi 1988.
A formal model of discourse structure, *Journal of Pragmatics* 12: 601–638.

12. Polanyi 1989a.
Telling the American Story. Cambridge, MA. MIT Press.

13. Polanyi 1989.
Clawe Permutation and Discourse Coherence. Talk presented to the Linguistic Society of America. Washington, D.C.

14. Polanyi 1991.
How to do things with Discourse. Talk presented to the Conference on Grammatical Foundations of Prosody and Discourse. Santa Cruz, CA.

15. Polanyi, L. and R. Scha. 1984.
A Syntactic Approach to Discourse Semantics in *Proceedings of the 6th International Conference on Computational Lingistics*. Stanford CA.

16. Prince, E. 1978.
A Comparison of WH–clefts and IT–clefts in discourse *Language* 54: 883–906.

17. Reichman, R. 1985.
Getting computers to talk like me and you. MIT Press.

18. Scha, R. and L. Polanyi. 1988.
An Augmented Context Free Grammar for Discourse in *Proceedings COLING88*. Budapest.

19. Scha, R., and H. Prust. 1990.
A Discourse Perspective on VP–Anaphora in M. Stokhof and L. Torenvliep (eds.) *Proceedings of the 7th Amsterdam Colloquium on Formal Semantics*. Amsterdam.

20. Webber, B. 1989.
Deictic Reference and Discourse Structure. University of Pennsylvania ms.

Related Disciplines – discourse analysis and use III

Rapporteur: Ellen Riloff

Bruce Britton:

Summarizing Situation Models: Using Principal Components to Reconstitute the Expert's Causal Model in the Reader's Mind

Bruce Britton proposed a text summarization model that uses principal components analysis to identify the main ideas in a text. A text is expressed as a correlation matrix where each entry represents the direction and magnitude of the relationship between two concepts in the original text. Principal components analysis is applied to the matrix to select the most useful concepts for a summary. The concept with the largest eigenvalue is the most useful concept because it can reconstitute the largest part of the matrix. In general, concepts are ordered in importance by their eigenvalues so they can be sequentially added to a summary to accommodate different length constraints.

Most of the discussion following this talk revolved around issues of text representation:

- Some people felt that it was unrealistic to assume a single representation of a text (e.g., correlation matrix) for different people. Britton responded that his model is for general text summarization and that we must assume some common associations among people. Experimental results show that, for expository texts, people generate knowledge structures that are highly correlated with one another. However, this does not hold true for literary texts.

- What is a "concept" in this model? What should the components represent? Can every word in the text be a concept? As a starting point, some people suggested that we could apply shallow processing techniques and use simple linguistic relationships as concepts.

- How do we obtain appropriate values for the correlation matrix?

- What is the relationship between world knowledge and knowledge in the text itself? Some of the associations in the matrix represent world knowledge (e.g., robbers & police) but others come from the content of the text (e.g., robbers & dead). How would you represent a conflict between these knowledge sources (e.g., if a policeman is a robber)? How would inferences be represented?

All correlations between concepts are represented as a number, but this does not capture different types of relationships between concepts. How would we represent different types of relationships? Is it sufficient to represent a text as a causal model?

Most people felt that it was impossible to judge the effectiveness of this approach without seeing concrete examples of how it would work with real texts.

Summarizing Situation Models: Using Principal Components to Reconstitute the Expert's Causal Model in the Reader's Mind

Bruce K. Britton

The method applies several techniques in sequence. It starts with an expert's causal model. Causal models can always be depicted as a diagram, composed of a set of variables whose names are printed in boxes, with the variables connected by causal arrows. Then this diagram is translated into its corresponding correlation matrix, using LISREL (Joreskog & Sorbom, 1990), taking advantage of the fact that correlation is essential for causality (i.e., only in the presence of a correlation between two things can one infer that one causes the other; Hume, 1748/1977). Next a principal components analysis of the correlation matrix is done using SAS or a connectionist network. This produces a set of principal components, each of which is a weighted linear combination of the variables. These principal components form the basis for writing the summary. That text is intended to reconstitute the experts' causal model in the novice's mind when the novice reads the text. Why do I think that a text based on principal components will reconstitute the experts' causal model in the novice's mind? The reason is that the mathematical meaning of principal components is that they are exactly those linear combinations of variables that will reconstitute the correlation matrix (Jolliffe, 1986). If the correlation matrix is the essence of causal knowledge, and if the readers can reconstitute the correlation matrix, then they have reconstituted the essence of the experts' causal knowledge. In writing the summaries, not all of the principal components will be used, but only a subset of them. My hypothesis is that we should present that subset which will allow the novice to reconstitute the experts' model to the maximum extent possible. That subset is the largest principal component. I base this hypothesis on the mathematical fact that the largest principal component will reconstitute more of its underlying correlation matrix than any of the other principal components. My claim is that the best possible way to communicate a correlation matrix in a limited time is first to describe the first principal component, then the second if there is time, and so on, to the last principal component if there is time to teach the whole model. If my method works, it will give us a way of summarizing causal models that is general, and also one that mathematical considerations suggest is optimal. It appears that procedural models can also be summarized in this way, because they are also expressible as correlation matrices.

Related Disciplines – discourse analysis and use IV

Rapporteur: Lucia Rino

Ines Busch–Lauer and Sumiko Mushakoji are both investigating summaries in the medical domain, particularly concerning author abstracts.

Their common and interesting conclusion is that author–produced abstracts do not always reflect the content of the corresponding article, even when they are written in the mother tongue of the writer.

Ines Busch–Lauer:

Abstracts in German Medical Journals – A Linguistic Analysis

According to Ines, the authors lack knowledge of discourse structure that results in missing information (for instance, missing conclusion or purpose of the investigation). They also rely very much on superficial constructions that make the resulting texts unclear and vague. The abstracts are often dull and do not convey the most relevant information of the corresponding article.

Although Ines' concerns are related to the ways a native English speaker perceives the message transmitted by a non–native writer, she notices that the problem of reception of information is related to the form and the extension to which a general reader feels motivated to read the corresponding paper.

She recognises that besides the linguistic view, other factors influence the production and reception of information. For example, there are presumptions about the audience that are not very clear in the abstracts she's analysed. Nevertheless, such presumptions are important in determining the relationship between linguistic competence of the writer and comprehension skills of the reader. Therefore, sociolinguistic factors must be taken into account.

Sumiko Mushakoji:

Constructing 'Identity' and 'Differences' in Original Scientific Texts and Their Summaries: Its Problems and Solutions

In Sumiko's approach, an abstract is the way two people communicate with each other. There is then a negotiation between the participants (i.e. the writer and the reader), in order to diminish differences and reinforce similarities during the process of comprehension. According to this view, authors do not refer only to the original texts to build their abstracts, but to the source of information from which the original texts have been built. Moreover, different instances of summaries of the same original text may be produced, depending on the situation during the process of communication.

Sumiko suggests then that such differences and similarities should be explored, not only in respect to theories of text structure, but also to sociological and cultural aspects. These aspects may lead not only to the selection of the most important components of the original text, but also to information that is relevant to the reader, but not apparent in the original text.

I see a common and important point in both presentations, related to the interplay between reader and writer, and source knowledge of both participants in the process of communication.

Constructing 'Identity' and 'Differences' in Original Scientific Texts and Their Summaries: Its Problems and Solutions

Sumiko Mushakoji

University of Library and Information Science, Japan

Original texts (hence, 'originals') and their summaries are often different and identifiable. There are, of course, many textual 'differences' between them. Yet it is common to recognize summaries as "condensed (information)" and to regard the two texts as the same or similar in terms of their information or content.

Although the relationship between originals and summaries exists in many kinds of texts, I will primarily focus on two kinds of scientific texts, i.e. original articles (hence, 'articles') and their author abstracts (hence, 'abstracts'). Abstracts are written by the very authors of the articles, and these authors have to maintain the same scientific information in two different texts.

My aim it to study how these differences and identity are constructed, and what the constructing identity and differences in scientific texts means to the actual process of transferring scientific information and handling scientific texts. Rethinking how far the present models can explain the problem of differences, which relates to producing ill or irrelevant abstracts, I take a more fundamental problem into consideration; the problem inherent in the relationship between a text and its source.

1. Text Summarization Model

The models of producing articles and their abstracts are almost entirely taken from the more general models of summarizing texts. Today the studies on summarization are already a rich research field, and many empirical studies gave us a new insight into the actual process of summarizing (Endres–Niggemeyer et al, 1991; Liddy, 1991; Kaplan, in press). Among these studies, now it seems that we can see some (theoretical) models and research frameworks in various fields and research domains. It is yet our consensus that "more powerful summarizing systems than those developed so far are clearly needed" (1).

The models of text summarization usually presume that the information or content of the originals are condensed into the summaries. Summaries are produced from the originals and help us to determine the information or content of the originals. These originals are, in the first place, produced from the source, i.e. some events, actions, or phenomena in the world. Summarization is

therefore the re–production process from the source.

This presumption permeates among our talks whenever we call a piece of text as "a summary" of another. We, as researchers and specialists in summarization and as more ordinary readers of summaries, assume that both the originals and summaries are based on the same source. This presumption illustrates a model of summarization as a linear one–way process:

```
< Source - Original (Production) - Summary (Reproduction) >
```

or two-forked processes from the same source:

```
< Source - Original (Full Text) - Summary (Smaller Text) >
```

In this study, I will call the model which explicitly or implicitly assumes either of these processes as 'Text Summarization Model'.

2. Problem of Textual Differences

Some empirical studies, such as discourse analysis, have shown the existence of unqualified abstracts (Salager–Meyer, 1991) and abstracts which do not reflect the macrostructure of articles (Milas–Bracovic, 1987). This inadequacy suggests that the differences between texts of articles and abstracts generate the problem that what articles are about and what abstracts are about do not correspond.

In order to clarify this problem, these differences are first of all to be analyzed. Seven categories of 'differences' have so far been discovered:

1. omission,
2. paraphrase,
3. arrangement,
4. supplement,
5. describing the visual information from charts, tables, and photos,
6. processing the cited information, and
7. formation of the characteristic styles of abstracts

(Mushakoji, 1988).

Comparing two kinds of texts, however, one cannot interpret their differences only from the surface structures of texts. For instance, one may be able to explain some omissions in the details of

articles, but it is impossible to find a good reason for deleting one experimental result and extracting another in abstracts. In order to interpret such differences, we must understand the source, i.e. what the author writes, and try to relate our understanding to the differences. Moreover, it is necessary to invoke not only the textual contexts but also more extensive sociological and cultural contexts of individual authors' writing. The simple linear proces
`< Source / Article - Abstract >`
is then in question.

3. Further Problem: an Asymmetric Relationship between a Source and its Text

This problem then leads to a further problem in the background assumption. When we transfer scientific information or present scientific knowledge, we do this by producing texts. That is, what we want others to know or to do is only recognizable by texts. Scientists as authors present their scientific activities and scientific findings as information by texts in a broad sense, and they formally present them by articles. Thus, generally, there exists an essential relationship between a text and its source.

This relationship is, however, an asymmetric one. It has been demonstrated in various ways, that when scientists' actions are accounted, there is a potential variability of scientists' statements about any given action. This has been a controversial topic in Sociology of Scientific Knowledge (SSK) since 1980s. Gilbert and Mulkay have analyzed extensive interview data and scientists' writings, and found that their discourses were

1. context–dependent, and
2. variable even in one talk by a single scientist.

This finding is problematic for us, too, because it questions our assumption about the relationship between a **source** and a **text** in general and thus the relationship between a **source** (of an article) and **its abstract** in particular. It is not clear how we can handle what is written in an article and what is written in its abstracts, nor whether we should assume one identical source in them. Furthermore, as a specialist of summarization, we should find out what authors do when they actually write the abstracts of articles.

4. Research on How Identity and Differences are constructed

Based on these problems, first of all, the textual differences when we treat articles and abstracts as two separate texts, were specified. The findings are to be combined with the ethnographic data, in order to explore how the identity is maintained in spite of the textual differences, in my future schedule.

4.1 Text Analysis

Data were taken from the articles written by biochemists at Tsukuba University who participated the finding of a vasoactive peptide, "endothelin". The articles published in 1988–89 were selected (20 articles). The texts of articles and abstracts were analyzed along the following aspects;

1. Semantic Analysis (propositional analysis)
2. Linguistic Analysis (modality, deixis, etc.)
3. Analysis of 'Style' (editing styles, technical jargon, etc.)

The titles of the articles were included in the analysis when necessary. During the analysis, I tried to make various tentative relations between the results and those of my previous studies (Mushakoji, 1988; Mushakoji and Nozoe, 1992; Mushakoji and Nozoe, 1993).

4.2 Ethnography

Collecting the ethnographic data, including non–standardized interviews and notes on participant observations, is under way. The data are also to be related to those of my previous study (Mushakoji, 1989).

5. Discussion and Some Solutions to the Problems

Examining abstracts and articles as two separate texts, their textual differences are clearly demonstrated. This finding questions us whether the sources, i.e. what an abstract is about and what an article is about, are identifiable or not. In my previous study, the authors never made a clear distinction between them, and legitimated the textual differences by their explanatory manipulations (Mushakoji, 1989). I presume that these differences are to be overcome by people (including scientists, laypersons, and us specialists in summarization) in their daily practices, so that they can identify the source of articles and abstracts. In other words, what makes the identity in the two different kinds of texts is in our continuous praxis of handling them as peculiar genres of texts. We recognize one piece of text as a summary and another as its original. We can, of course, claim that this piece of text is an irrelevant or ill summary, but we can do this because we do recognize it as "a summary". We are constructing its identity to the original.

On the other hand, this kind of construction is dynamic. The authors accounted for their abstracts in a variable way (Mushakoji, 1989). There was a considerable confusion about the source, articles, and abstracts when the authors accounted for their abstracts. For instance, to the question about abstracts in the previous interview, the authors' responses sometimes consisted of the answers

about articles or even about their research itself. Whether abstracts are produced from the articles or from the source, was often variable. This is supported by the finding in the Text Analysis which indicated that the texts of abstracts were either located on the articles or on the researches in the deictic context.

Confronted with the problem of textual differences in articles and abstracts, and the further problem of the aymmetric relationship, what will be our solution? Here are four (tentative) ones drawn in this study;

1. **Solvable**; by finding the rules or theories in text structures
2. **Solvable**; by finding the rules or theories in human cognitive process, social context, or cultural background of producing summaries
3. **Unsolvable**; let's admit that we cannot conquer these 'differences'. Summaries are totally different texts whose tie to the originals is socially negotiated. They are not smaller copies / miniatures. It is impossible to condense information.
4. **(Un?)solvable**; we should rather celebrate these 'differences'

The source of originals and those of summaries, i.e. "scientific facts", "nature", or "scientific phenomena", can only be presented in texts that have unavoidable variability. It is not escapable that "recall and summarization include interpretation" when we use language in interactional settings (Cicourel, 1978).

But on the other hand, when we participate in the flow of scientific information, or the process of knowledge accumulation, we are unavoidably facing 'identifiability' of the source, which are to exist in the series of producing different texts. In this series, the expression of each text is not the same at all, but it acts as a device that tells us what the previous texts **have been** about at this point of time. The expression

> *"with this experimental material (say, the left carotid artery of male Wister–Kyoto rats) catheterized in the X, ... the arterial blood pressure was decreased within 5 hours. ... These data suggest so–and–so. ... and those data suggest so–and–so. ... Therefore, the results of this study indicate that X is efficacious for hypertension",*

would be processed successively in a series of various scientific texts by the author and other researchers, and may become just

> *"X is efficacious for hypertension" .*

Textual processes such as generalize, simplify and summarize are to create another text, which contains the different statement about that source, and to add a new text to the end of the series,

under the name of one identifiable source. That is, we re–state the original statement and we construct the identity in our current accounts. And by this process, we are making sense of transferring scientific information, and participating in the settlement of this statement into knowledge or ignorance. Summarization is no exception in this series of constructing identity and differences.

Footnote

1)

Draft for "Summarising Text for Intelligent Communication, Dec. 12–17, 1993, Dagstuhl seminar 9350".

References

Cicourel, Aaron V. (1978)

"7 Interpretation and summarisation: issues in the child's acquisition of social structure". The Development of Social Understanding. New York, Gardner Press. p.251–281.

Endres–Niggemeyer, Brigitte et al. (1991)

Modelling summary writing by introspection: a small–scale demonstrative study. TEXT, Vol.11, No.4, p.523–552.

Gilbert, G. Nigel and Mulkay, Michael (1984)

Opening Pandora's Box: a Sociological Analysis of Scientists' Discourse. Cambridge, Cambridge University Press.

Kaplan, Robert B. et al. (in press)

On abstract writing. TEXT.

Liddy, Elizabeth DuRoss (1991)

The discourse–level structure of empirical abstracts: an exploratory study. Information Processing & Management, Vol.27, No.1, p.55–81.

Milas–Bracovic, Milica (1987)

Structure of scientific papers and their author abstracts. Informatologia Yugoslavica, Vol.19, No.1–2, p.51–67.

Mushakoji, Sumiko (1989)

The process of forming information into media: qualitative analysis centered upon author abstracts. Library and Information Science, No.27, p.15–35.

Mushakoji, Sumiko (1988)

Qualitative analysis for the relation between original articles and their abstracts: an approach toward the condensation of information. Library and Information Science, No.26, p.1–29.

Mushakoji, Sumiko and Nozoe, Atsutake (1992)

Discourse analytic approach toward scientific literature: centering upon <OMISSION> in author abstracts from the Introduction Parts of their original articles in medical science. Research Report of University of Library and Information Science, Vol.12, No.1, p.1–27.

Salager–Meyer, Francoise (1991)

Discoursal flaws in Medical English abstracts: a genre analysis per research– and text–type. TEXT, Vol.10, No.4, p.365–384.

Abstracts in German Medical Journals – A Linguistic Analysis

Ines Busch–Lauer

Leipzig University

Studies on text types and contrastive genre analysis have become current issues in LSP research and are intended to economize specialist communication.

Abstracts are one of the most widely used research–process genres (Salager–Meyer 1990; Staheli 1986; Swales 1990). According to the ISO 214 – 1976 (E) the term abstract "signifies an abbreviated, accurate representation of the contents of a document, without added interpretation or criticism and without distinction as to who wrote the abstract".

Although there exists much general advice on abstract writing in native speaker (NS) handbooks and style manuals of English there is still a lack of published work that offers real and specific help to the non–native speaker (NNS)/writer of English.

The present paper briefly introduces the aims and methodology of a wider research project investigating cultural and cross–linguistic similarities/differences in written academic discourse (both English and German) in the field of medicine covering the point of view of both LSP workers and specialists in the subject area, i.e. medical researchers, doctors and students. To find out contrasts in the discoursal structure of special texts and their linguistic manifestation various research–process genres (research articles, abstracts, argumentative and descriptive essays, research presentations, review articles) will be studied according to a complex linguistic top–down approach. Questionnaires and the 'thinking aloud' method will be applied to study the writing process of NS and NNS of English and to mark different thought patterns.

The major part of the paper discusses results of the linguistic analysis of German and English abstracts in German medical journals. Teaching experience shows that writing is the most challenging and creative aspect of medical communication. It is a false but widely held assumption that if a researcher can write a coherent research article in the mother tongue it will not cause any trouble to submit an abstract of 150 – 200 words in a foreign language as often required by German medical periodicals. Due to constraints of time and a more or less 'fossilized' linguistic competence of English many German researchers feel unequal to this task and basically rely on translation services which often fail to produce both accuracy regarding the medical contents and linguistic appropriacy in their piece of information.

But titles and abstracts in published papers are both front and summary matter of research and as they are often the only available and read source of information they should be self–explanatory in contents and neat in structure and shape.

To find out the formal schemata and linguistic devices of German abstracts and their NNS English equivalents opposed to findings on NS abstracts (Salager–Meyer 1990) a corpus of 20 abstracts taken from German research journals of various degrees of specialization was studied according to the following criteria: overall length (sentences, words); length of moves (sentences, words); macrostructure; moves; linguistic devices (linking words/connectors, term density, tense, verbs); information content in comparison with the reference text.

The analysis shows that German medical abstracts are often not very well– structured. The moves described by Salager–Meyer (1990) are only partly identifiable in the corpus. Some abstracts are too wordy and do not represent the structure and overall research results. Extra–ordinary long sentences hamper readability and comprehension. A comparison between German abstracts and their English equivalents indicates that authors largely rely on their own ability to translate texts into English which transfers academic writing traditions into an 'artificial' English conveying wrong pieces of information.

As a conclusion of this study the author suggests a checklist for German students and specialists for composing a well–structured abstract to be discussed by the audience. Furthermore there is the need of further research in the subject area to confirm the preliminary results and to implement special language courses for students and scientists to train the art of writing, summarizing and abstracting.

References

ISO 214–1976 (E). International Organization for Standardization:
Documentation: Abstracts for publication and documentation., 1976. Genf.

Salager–Meyer, F. (1990):
Discoursal movements in medical English abstracts and their linguistic exponents: a genre analysis study. *Interface*. Journal of Applied Linguistics 4.2., 107–124.

Salager–Meyer, F. (1992):
A text–type and move analysis study of verb tense and modality distribution in medical English abstracts. *English for Specific Purposes*. An International Journal Vol. 11, Number 2, 93–113.

Staheli, L.T. (1986):
Speaking and writing for the physician. New York: Raven Press.

Swales, J.M. (1990):
Genre analysis. English in academic and research settings. Cambridge, UK: Cambridge University Press.

Computational Resources I

Rapporteur: Mark Maybury

This theme focuses on how linguistic and extra–linguistic knowledge and processes might be applied to summarization for multiple purposes (e.g., find, extract, generalize information).

Parsing, Linguistic Resources & Semantic Analysis for Abstracting and Categorization

William Black

In his talk William described linguistic resources for automatic abstracting and text categorization. For abstracting, he describes the indicator phrase (IP) method (Paice, 1982) which simply uses weighted rules that detect IPs (e.g., "the objective of this", "the primary point is") to determine what to included in a resulting abstract. He found two weaknesses with this approach: loss of coherence and cohesion. He, in part, addressed the latter by deleting dangling "it" anaphors using a small rule base that distinguishes anaphoric vs. non–anaphoric "it" in 90% of cases. If the preceding sentence was deemed necessary to resolve an anaphor, it was included. As definite noun phrases can be referential too (e.g., "the experiment"), a subsequent project, BLAB, eliminated these as well. Basically, BLAB abstracts by

1) retaining sentences using IP metrics and
2) rejecting unresolvable anaphoric sentences using (a set of 9) syntactically oriented rules (e.g., reject if subject is anaphoric pronoun as object pronouns are assumed to be resolved in same sentence).

While this method is crude, like Dialog's lead–line abstractor, and does not work all the time (~80%?), it seems effective and improvable by at least 30% (although Karen Sparck Jones was less optimistic about ease of improvement, citing language processing experiences). Black felt this approach would work well on biography and shootout texts, although it elided imperative sentences.

In the second half of his talk, William described the COBALT project, which focuses on categorizing news wire texts for market specialists. He contrasted their linguistically–motivated approach with the commercial text categorization shell (TCS) based on text patterns (e.g., skip # words) and domain specific categorization rules, indicating the importance of, e.g., predicate/argument structure. He noted linguistic resource requirements for categorization,

indicating it must be: robust, bottom–up (to deal with unseen texts); have an extensible lexicon (he cited the 1.5 million different proper names in US telephone directory), produce a single analysis (underspecification ok), produce a predicate–argument structures, and be fast.

He concluded by indicating results from an informal evaluation of sentence extraction methods (IP and word frequency based techniques) versus human abstracts. He found subjects could distinguish between the methods, although they said content indicativeness was roughly equivalent, human extracts were more cohesive, and the IP summaries were more readable than the statistical ones.

Jerry Hobbes commented that given heuristics like first sentences have no anaphors, techniques even simpler than rules (e.g., choosing the first X characters) should yield a good summaries.

Generating the Complex Sentences of Summaries using Syntactic and Lexical Contraints: Two Applications

Kathy McKeown

Kathy described work in collaboration with Jacques Robin (Columbia University) and Karen Kukich (BellCore) which focuses on generating the text of a summary as opposed to determining the information to be included from an article. McKeown described the salient features of the work as generating genre and domain specific, informative summaries (in place of full text) from input data (vs. text). She commented that these combined features make this approach feasible. She took a descriptive as opposed to prescriptive methodology, by analyzing human summaries to discover summarization devices. She described two applications. The first summarized quantitative (in this case, basketball) scores by summarize highlights and underscoring historical significance. The second application generated summary and detailed documentation of telephone planner activities using a trace of their interaction with a software planning system.

The key problems addressed by both applications were what information to include in the summary and how to pack information in limited space using lexico–syntactic devices. She characterized her summaries (source and generated?) as having long sentences (21–46 words in length) which loaded information using multiple modifiers, conjunction, ellipsis, and that added information opportunistically. She distinguished between mandatory (main points, fixed, expected) and optional (floating, historical significance) information. (What about unexpected information?). McKeown indicated her summaries were genre–specific, using long sentences and active verbs.

Kathy contrasted this work with previous report generation systems (ANA and FoG), which did not include optional (it floats, hence hard) information, which was nevertheless significant both

quantitatively (40% lead sentences included floating information) and qualitatively (historically significant information always appears as floating). She then went on to outline Jacques Robin's (Columbia, Kathy's student) partially–implemented generation architecture based on a draft & revision approach (to deal with floating and sentence complexity) guided by a corpus analysis of 300 lead sentences from 800 basketball reports. This analysis identified 4 fixed information units (result, score, date, most notable final statistic from the winning team player) and 3 predominate floating types (final statistics, records, streak). Moreover, revision tools identified included 4 simple mechanisms (e.g., single attachment – adjoin, absorb) and 6 complex (transform structure – adjunction, nominalization (e.g., to add adjective)), which have been partially implemented.

The second system, PLANDoc (with Karen Kukich), generates English narratives which document telephone network planning from use of the LEIS–PM system (documentation for managers & regulators, 1500 engineers, 8 BellCore Companies, 10–15 routes/years, 15,000 documentation packages/year). Notably, the current documentation structure (Table of input, system summary plan, alternatives, summary recommendation) includes two imbedded summaries. Research issues include what to include, where to include it, that fact that the summary is not a direct translation (information is added, e.g., because it fits).

Kathy concluded by indicating that generating concise summaries is possible, lexical/syntactic constraints control how information is selected, this work generates incrementally, and address the issue of mandatory and optional information.

Applying her incremental, revisional techniques to the shootout example text yielded a felicitous result, in terms of content captured, readability, and space required.

Generating the Complex Sentences of Summaries Using Syntactic and Lexical Constraints: Two Applications

Kathleen McKeown (1)

Department of Computer Science
450 Computer Science Building
Columbia University
New York, N.Y. 10027

1 Introduction

Like machine translation, summarization is a task that seems to require solutions in both interpretation and generation. Summarizing an article, for example, involves interpretation of the article to identify the most important facts and generation to produce a paragraph that conveys those facts concisely. Contrary to this view, we have identified several summarization tasks involving generation only. We are working in two domains, summarization of sports reports [Robin and McKeown 93, Robin 93] and automated documentation of telephone planning engineer activities [Kukich et al. 93]. In the first of these domains, input is a set of box scores for a basketball game and the task for the system is to summarize the highlights of the game, underscoring their significance in the light of previous games. In the second domain, the system must produce a report documenting how an engineer investigated what new technology is needed in a telephone route through use of a sophisticated software planning system, `LEIS-PLAN`. Input to the generation system is a trace of user interaction with `LEIS-PLAN` and output is a 1–2 page report, including a paragraph summary of activity as well as a detailed narrative.

In both of these domains, the problems for summary generation are the same:

- What information should be included in the summary
- How to pack in as much information as possible in as short an amount of space as possible
- How to use syntactic and lexical devices to convey information concisely
- Given the choice of a single word or syntactic structure, how does this constrain (or allow) the attachment of additional information

Through analysis of example summaries in both domains, we have found that summaries typically use quite complex sentence structure to load maximal information into a single sentence. Summary sentences most often include multiple modifiers, conjunction, and ellipsis. Furthermore, it appears that some information is opportunistically added into the summary, based on words and syntactic

structure used to realize mandatory information. These observations provide some answers to the questions above; they suggest that there is a set of information which much be included in the summary and that secondary information is included when the form of a base summary allows. Furthermore, the syntactic structure and words of the base summary provide hooks onto which additional optional information can be added.

In the following sections, we describe each application domain, showing how summaries can be generated from structured data and the complexity of the sentences. We also show specific syntactic and lexical constraints for the two domains, which both make it possible to generate the complex sentences that are required and allow some information to be included based on surface features.

2 Generation of Summaries of Quantitative Data

We are developing a system to generate summaries similar to newswire basketball game summaries [Robin 93, Robin and McKeown 93]. Input will be a set of box score statistics for a given game. The system will ultimately consist of three components, a content planner, reviser and sentence generator. The content planner will first select mandatory information from the box scores that should be included in the summary. Our analysis of newswire basketball summaries shows that information such as teams, who won and lost, and significant statistic of one player consistently appear in lead sentences of reports. This will be used to form a "draft". Next, the content planner will use a history database of box scores from previous games to determine which historical statistics the current facts stand in contrast to. For example, if a player's score is the highest score of any player over the season this fact should be noted. A reviser and sentence generator, currently being implemented in STREAK [Robin 93] take the facts for the draft, produce a text and then use constraints from the draft and the semantics of optional new facts to determine how to fold in additional historical information. The content planner is in design stages, while our focus to date has been on the development of the reviser and sentence generator.

Summarization of quantitative data raises several challenges for language generation systems. First, sentences in such reports are very complex (e.g., in newswire basketball game summaries the lead sentence ranges from 21 to 46 words in length). Second, while some content units consistently appear in *fixed* locations across reports (e.g., game results are always conveyed in the lead sentence), others *float*, appearing anywhere in a report and at different linguistic ranks within a given sentence. Floating content units appear to be opportunistically placed where the form of the surrounding text allows. For example, in examples 1–3 below, sentences 2 and 3 result from adding the same streak information (i.e., data about a series of similar outcomes) to sentence 1 using different syntactic categories at distinct structural levels. Although optional in any given sentence, floating content units cannot be ignored. In our domain, they account for over 40% of lead sentence

content, with information conveying the historical significance of facts *only* conveyed as floating structures. Most report generators to date [Kukich 83, Bourbeau et al. 90], however, cannot include historical information precisely because of the fact that it floats.

1. *Draft sentence:* "San Antonio, TX — David Robinson scored 32 points Friday night lifting the San Antonio Spurs to a 127 111 victory over the Denver Nug gets."
2. *Clause coordination with reference adjustment:* "San Antonio, TX — David Robinson scored 32 points Friday night *lifting the San Antonio Spurs to a 127 111 victory over Denver* **and handing the Nuggets their seventh straight loss**".
3. *Embedded nominal apposition:* "San Antonio, TX — David Robinson scored 32 points Friday night lifting the San Antonio Spurs to a 127 111 victory over *the Denver Nuggets*, **losers of seven in a row**".

To determine how floating content units can be incorporated in a draft, we analyzed a corpus of basketball reports, pairing sentences that differ semantically by a single floating content unit and identifying the minimal syntactic transformation between them. The result is a set of revision tools, specifying precise semantic and syntactic constraints on (1) where a particular type of floating content can be added in a draft and (2) what linguistic constructs can be used for the addition.

We distinguished two kinds of revision tools. Simple revisions consist of a single transformation which preserves the sentence of the draft, adding in a new constituent. Complex revisions, in contrast, are non–monotonic; an introductory transformation changes the draft sentence in adding in new content, often resulting in an ungrammatical sentence. Subsequent restructuring transformations are then necessary to restore grammaticality. Simple revisions can be viewed as elaborations while complex revisions require true revision.

Adjoin is one example of a simple revision tool that can be used to insert additional constituents of various syntactic categories at various syntactic ranks. It adds an adjunct under a constituent head in the draft sentence. Sentence 3 above shows an example of nominal rank adjoin of an appositive noun phrase. **Nominalization** is an example of a complex revision tool. It replaces a meaning carrying verb in the draft sentence, with a collocation consisting of a support verb plus a noun, where the noun is a nominalization of the original verb. This type of revision allows additional modifiers to be added onto the nominalization in subsequent revisions. For example, in revising sentence 4 into sentence 5 in Figure 2 below, the verb structure "X defeated Y" is replaced by the collocation "X handed Y a defeat." Once nominalized, "defeat" can then be pre–modified by the phrase "their sixth straight home" providing historical background. This rule embodies both a syntactic and lexical constraint on adding in new information modifying an action; if the verb describing the action can be nominalized, then the information is selected.

This approach to generation allows STREAK to incrementally produce the complex sentences required for summarization by opportunistically adding in optional historical information as the syntactic structure and semantics of the draft allows. An example of how this would work is shown below in Figure 2 where five different revision tools are applied in sequence, each adding in new information. This example is partially implemented. We are currently carrying out an analysis of another quantitative domain, stock market reports, to determine the domain independence of the revision tools we identified.

---

1. **Initial draft (basic sentence pattern)**: "Hartford, CT — Karl Malone scored 39 points Friday night as the Utah Jazz defeated the Boston Celtics 118 94."
2. **adjunctization**: "Hartford, CT — Karl Malone **tied a season high with** *39 points* Friday night as the Utah Jazz defeated the Boston Celtics 118 94."
3. **conjoin**: "Hartford, CT — Karl Malone tied a season high with 39 points **and Jay Humphries added 24** Friday night as the Utah Jazz defeated the Boston Celtics 118 94."
4. **absorb**: "Hartford, CT — Karl Malone tied a season high with 39 points and Jay Humphries **came off the bench** *to add 24* Friday night as the Utah Jazz defeated the Boston Celtics 118 94."
5. **nominalization**: "Hartford, CT — Karl Malone tied a season high with 39 points and Jay Humphries came off the bench to add 24 Friday night as the Utah Jazz **handed** the Boston Celtics their sixth straight home *defeat* 118 94."
6. **adjoin**: "Hartford, CT — Karl Malone tied a season high with 39 points and Jay Humphries came off the bench to add 24 Friday night as the Utah Jazz handed the Boston Celtics their **franchise record** sixth straight home defeat 118 94."

Figure 1: Incremental generation of a complex sentence using various revision tools

---

3 Automated Documentation of Planning Engineer Activity

Jointly with Bellcore (2), we are developing a system, PLANDoc, that will document the activity of planning engineers as they study telephone routes. The telephone network planning engineer's job is to derive a capacity expansion (relief) plan specifying when, where, and how much new copper, fiber, multiplexing and other equipment to install in the local network to avoid facilities exhaustion. Planning engineers currently use a software tool, the Bellcore LEIS-PLAN system, that helps them derive a 20–year plan based on economic and usage constraints. Documentation of the activity is helpful for informing managers who are responsible for authorizing expenditures as well as for auditors and external regulators.

Input to the generation system, PLANDoc, is a trace of the engineer's activity with LEIS-PLAN. It indicates, in table format, the different types of changes to the route the engineer experimented with and how this affects overall cost. We based development of PLANDoc on a user–needs analysis (see [Kukich et al. 93]), including a set of model narratives written by an experienced planning engineer. This corpus provided a general model for the documentation, beginning with a summary of the plan produced by LEIS-PLAN alone, followed by a narrative describing the engineer's refinements to this plan, and concluding with a summary of the plan recommended by the engineer (which usually includes elements of the program generated plan and human refinements). Currently, PLANDoc can produce a narrative of the refinements (an example is shown below in Figure 2) and we are beginning work on generation of the summaries.

One clear issue in the organization of the summaries is the need for decisions about what information should be included in the summary to be influenced by how information can be grouped together in a single sentence. In order to convey information more concisely, similar refinements are often grouped together and short phrases that can refer to the group selected, usually using conjunction and ellipsis. Thus, for example, in one of our model summaries shown below in Figure 3 , the summary groups together a set of refinements which activated a specific type of equipment and uses a nominalization to refer to them ("DLC activations"). The order in which these refinements were actually carried out may be quite different and they may be separated by other actions in the input trace. However, the fact that they can be grouped together using a simple syntactic device such as conjunction and ellipsis and referred to concisely using nominalization determines its selection and its ordering in the summary.

RUNID Reg1: This refinement activated CSA's 3122, 3130, 3134, 3208 and 3420 for DLC in the third quarter of 1994. DLC system idlc272 was used for all placements in CSA 3122. For this refinement, the resulting 20–year route PWE was $2110K, a $198K savings over the base plan, and the 5–year IFC was $1064K, a $65K penalty over the base plan.

Figure 2: System Generated Narrative

The refinement "fib_2551" is the best solution for this route. This plan includes the BASE plan DLC activations of CSA 2907 in 1997 and CSA 2119 in 1998. It also includes the BASE plan fiber activation of CSA 2317 in 1994Q4.

Figure 3: Portion of a Model Summary

---

4 Conclusion

In both domains, the examples illustrate that selection of information to include in a summary can depend on lexical and syntactic constraints. The revision rules of STREAK, for example, can be triggered by a verb which can be nominalized, while in the PLANDOC domain, the choice to include one piece of information automatically triggers the addition of others if they can be concisely conjoined to the first using ellipsis. This approach of incrementally adding in information as it fits makes it possible for a generator to produce the complex sentences required in summaries. Finally, both applications illustrate the feasibility of summarizing input data as opposed to full text, making the overall summarization task tractable.

Footnotes

(1) This paper is based on work being carried out jointly with Jacques Robin, in the case of STREAK, and Karen Kukich, James Shaw, Jacques Robin, and Jong Lim, in the case of PLANDOC. STREAK is being implemented by Jacques Robin as part of his dissertation work.

(2) PLANDOC is being developed collaboratively with Karen Kukich and Neal Morgan of Bellcore and James Shaw, Jacques Robin, and Jong Lim of Columbia Univ.

References

[Bourbeau et al. 90]

Bourbeau, L. and Carcagno, D. and Goldberg, E. and Kittredge, R. and A. Polguere. Bilingual generation of weather forecasts in an operations environment. In *Proceedings of the 13th International Conference on Computational Linguistics*. COLING, 1990.

[Kukich 83]

Kukich, K.
Design of a Knowledge Based Text Generator. In *Proceedings of the 21st ACL Conference*. ACL, 1983.

[Kukich et al. 93]

Kukich, K., McKeown, K., Morgan, N., Phillips, J., Robin, J., Shaw, J., and Lim, J. User–Needs Analysis and Design Methodology for an Automated Documentation Generator. In *Proceedings of the Bellcore/BCC Symposium on User–Centered Design: "People and Technology"*. Piscataway, N.J., October, 1993.

[Robin 93]

Robin, J.
A Revision–Based Generation Architecture for Reporting Facts in their Historical Context. In Horacek, H. and Zock, M. (editor), *New Concepts in Natural Language Generation:*

Planning, Realization and Systems. Frances Pinter, London and New York, 1993.

[Robin and McKeown 93]
Robin, J. and McKeown, K.R.
Corpus Analysis for Revision–Based Generation of Complex Sentences. In *Proceedings of the National Conference on Artificial Intelligence*. Washington, D.C., July, 1993.

Parsing, Linguistic Resources and Semantic Analysis for Abstracting and Categorisation

William J. Black

Centre for Computational Linguistics
UMIST, Manchester

January 17th, 1994

(*)

1 Introduction

This paper discusses the linguistic aspects of text processing, based on experience with two related applications: abstracting of technical papers and text categorisation. Whereas systems addressing both applications can be and have been developed without formal linguistic analysis, we have preferred to develop an architecture in which re–usable linguistic resources and analysers play a part.

2 Abstracting by extracting

The extraction approach to abstracting is one which seeks to achieve robustness in simplicity, acknowledging the prematurity of natural language processing for such tasks. Essentially, an abstract is made up by concatenating sentences extracted from the source text by a mechanism that selects them as content–indicative. That mechanism can be based on pattern–matching, as described by Paice (1981) and Black and Johnson (1988), or it can be statistical, as described by Luhn (1958), Edmonson (1969) or Earl (1970). However, merely concatenating text sentences together risks incoherence, particularly noticeable if the extraction criteria take no account of chains of reference in the texts. A graphic example encountered by Black and Johnson (op cit) was a paper describing a sequence of three experiments, where three sentences were extracted each containing "the experiment". Each such token referred to a different experiment. In this case, the presence of anaphora in the extract renders it insufficiently coherence for use as an abstract. Liddy et al (1987) also report on the effect of the occurrence of anaphora on the statistical base for numerical measures of concept occurrence. This time, the presence of anaphora has a deleterious effect on the selection part of the extracting process. For both reasons, an important refinement of extraction–based abstracting is to attempt to control for the use of pronominal anaphora and other referring expressions.

Paice and Husk (1987) reported a relatively small rulebase which discriminates referring from non–referring uses of the pronoun "it" with a high degree of accuracy, and Liddy et al (op cit) six rules to do the same for "that". Armed with such rules, an extracting program can assimilate sentences preceding those in which a referring pronoun occurs. Pronominal anaphora do not, with very rare exceptions, refer further back in texts than the preceding sentence. However, the situation is much less straightforward when non–pronominal reference is taken into account.

The *BLAB* project was constituted to study the extracting process in such a way that definite noun–phrase referring expressions could occur in extracts without undermining coherence. An alternative method of extracting was developed which used logical aspects of discourse structure as a theoretical basis. Sentences containing referring expressions cannot be interpreted independently. Another way of saying this is that they are not propositional. In the same spirit as the earlier projects, *BLAB* did not seek to *resolve* the referring expressions, but to discriminate between occurrences of propositional from non–propositional sentences. Details of this approach are given in Paice et al (1993) and Johnson et al (1993).

A relatively small set of rules seemed to be effective in discriminating between referring and non–referring uses of "the". In implementation, these rules were reduced to 6 in total. This produced an alternative to the selection–based method used in the previous work, and was instead based on elimination of sentences that would, taken on their own, render the extract incoherent. This produced extracts that were much longer than those produced under the selection methods, containing around 20% of the original. The method provided no basis for tailoring the length of the extract (unlike the indicator–phrase selection method which weighted and ranked sentences for their content–indicativeness).

In this work, some evaluation methods were developed for comparing content–indicativeness with expert–tagged extracts, and for evaluating coherence. Description of these methods and results are in Paice et al (1993) and in a paper in preparation.

Another outcome was that the extracting and coherence–preserving rules were implemented within a modular architecture, and could be interfaced to different preprocessing and surface analysis components. This allowed evaluation to be conducted on different components separately.

2.1 Linguistic resources for abstracting

At the outset of the project, we did not know what the solution to the problem posed by definite referring expressions would be, and hence what input data the rules would work on. But it was suspected that a more sophisticated linguistic analysis might be needed than had been the case for

pronominal anaphora. In any case, a new team would work on this problem and the previous specially–developed rule language, "GARP" was not thought easily maintainable. Thus a surface linguistic analyser was developed that would provide linguistic descriptions on which discourse–level rules could operate.

The syntactic analyser for such a system must be first and foremost robust. It should produce some result on any input. This seems to force some basic choices. For example, bottom–up processing is to be preferred, although it is also a requirement that the analyser can abduce the syntactic description of much of the vocabulary used from the local syntactic context. Another requirement is that it should not produce too many analyses. One is ideal. A syntactic level of description at least avoids the generation of some quantifier scope ambiguities, but there are many potential structural ambiguities that one might take into account. The solution taken in *BLAB* to the problem of prepositional phrase attachments was not to define a rule which could attach them to an antecedent. Sub–second "parsing" is then possible for quite long sentences. A second set of attachment heuristics operates after the initial parse to deal with such problems (1).

There is certainly a limit to the potential for purely structural approaches to abstracting, as exemplified by *BLAB*, but the problems of discourse structure that this work raised are also relevant to approaches depending more on subject knowledge. However, the linguistic resources needed for a more truly knowledge–based approach to abstracting would almost certainly have to permit lexical semantic analysis as well as structural anaylsis. This is also true of the new application in which we are working on text processing, namely categorisation.

3 Categorisation

Text categorisation is more like indexing for current awareness than abstracting. The task is to assign texts from a source such as a newswire to categories related to the job functions or interests of the members of the consuming organisation. Nevertheless, many of the characteristics of the application are the same. Robustness of analysis and speed are probably as important as richness of analysis at least for the time being. On the basis of a linguistic analysis and the semantic processing needed for categorisation, it should be possible to provide a text summarisation service by analysis and generation. Nonetheless, one clear difference between abstracting as done by *BLAB* and categorisation as being done by *COBALT* is that in the latter case, semantic processing is of the essence.

The *COBALT* project is based on the adaptation and integration of components from previous text–processing projects. For surface syntactic analysis, which we describe below, the basis can be traced from the *BLAB* analyser through the linguistic resources developed for generation in a dialogue project, *PLUS*. For the semantic analysis, the antecedent is based on the *NOMOS* project,

whose objective was knowledge–base construction via text analysis. In *NOMOS*, semantic analysis involved a series of processes driven by heuristic rules, operating on syntactic trees to do conceptual disambiguation, collation of analysis fragments, resolution of attachment ambiguities etc. It was however related to a quite different text genre, namely legislative texts, from the newswire data for *COBALT*, and this imposes at the very least changes to the content of the semantic heuristics.

3.1 Linguistic Resources for Categorisation

It is possible to approach categorisation in such a way that linguistic analysis is hardly required, or at least it need not be based on conventional syntactic analysis. The *CONSTRUE* application developed by Reuters and Carnegie Group uses a facilitating software shell known as TCS. This embodies a pattern–matching language in which parts of the syntactic context are represented by gaps specified only by length between pairs of words assumed to be in semantic relation. This is very much the same approach taken earlier to extracting by pattern–matching (Paice, 1981), and it can also be seen to have some of the characteristics of the approach to NLP described as "semantic grammar". It is our contention that such an approach has several defects, despite an initially impressive performance on unseen texts. For one thing, whilst an "amateur linguist" can develop a TCS rule–base, the non–linguistic approach fails to capture generalisations, is bound to produce lower precision than equivalent semantic discrimination rules operating on a basis of analysed test, provides little in the way of knowledge engineering methodological support, requires more to be redone in porting the generic application to new concrete cases, and is incapable of sustaining an evolution of the application requirements beyond simple categorisation. Of course, these contentions remain just that at present, since this is ongoing work, so the remainder of this abstract describes the approach being taken to linguistic processing for categorisation and the rationale for various choices that have been made.

3.2 The *COBALT* Linguistic Analysis Module

Like the *BLAB* analyser, this is bottom up. Like its antecedent in *PLUS*, it produces a quasi–logical form (2) as output. This seems to us to be a reasonable level of initial description for genuine lexical semantic analysis for several reasons: It is easier to specify the interface with the semantic component independent of the linguistic resources to be used than would be a syntactic tree, whose topography is based on the linguistic rules used; It is, however, a better alternative than a fully–scoped logical form, since it is easier to minimise the number of competing analyses at this level; It may never be important for this application to resolve quantifier scopes.

The linguistic analysis used is a unification–based categorial grammar (3), augmented by function composition and type–raising rules (and supported by derivational equivalence–based methods for

eliminating spurious ambiguity, as described by Barry (1988) and Hepple and Morrill (1989)). Within this framework, it is easy to experiment with different approaches to such important choices as whether prepositional phrases subcategorise for their attachments or whether heads subcategorise for their optional modifiers. This is very much current work and will be elaborated on in the presentation.

The linguistic resources are being developed with the aid of a corpus of newswire texts that has been made available to us, and which is also being used for semantic rule development and application based categorisation rules. Special effort is being directed to the analysis of proper names and other "sublanguage" features. The design of the analyser intentionally assumes an incomplete lexicon.

Footnotes

*In automatic abstracting research, we have been collaborating with Paice at Lancaster on refining extraction–based approaches by taking more account of linguistic discourse structure. This has been done in the context of a British Library–funded project *BLAB*. In categorisation, we are working with two AI–oriented companies, Quinary SpA of Milan and Step Informatique of Paris, in a CEC–funded project *COBALT* within the Linguistic Research and Engineering programme. The author gratefully acknowledges the financial support of the two grant–awarding bodies and Brother International PLC.

1) to the extent that it is necessary to deal with such ambiguities, since the referring/non–referring discrimination does not often depend on postmodifiers except for "of".

2) Quasi–logical form is intentionally indefinite here.

3) Again, indefiniteness is intentional.

References

Barry, G.D. (1988)
Parsing Strategies for Categorial Grammars. MSc Dissertation, University of Manchester.

Black, W.J. and Johnson, F.C. (1988)
A Practical Evaluation of Two Rule–Based Automatic Abstracting Techniques. *Expert Systems for Information Management* **1**(3), 159–177.

Earl, L.L. (1970)

Experiments in Automatic Abstracting and Indexing. *Information Storage and Retrieval* **6** (4), 313–334.

Edmonson, H.P. (1969)

New Methods in Automatic Extracting. *JACM* **16** (2), 264–285.

Hepple, M. and Morrill, G. (1989)

Parsing and Derivational Equivalence. Proc. 4th EACL, Manchester, 10–18.

Johnson, F.C., Paice, C.D., Black, W.J. and Neal, A.P. (1993)

The application of linguistic processing to automatic abstract generation. *Journal of Document and Text Management* **1**(3) *to appear.*

Liddy, E.D. et al. (1987)

Liddy, E.D. (reference is missing here).

Luhn, H.P. (1958)

The Automatic Creation of Literature Abstracts. *IBM J. R & D* **2**(2), 156–165.

Paice, C.D. (1981).

The Automatic Generation of Literature Abstracts: An Approach Based on the Identification of Self–Indicating Phrases. *In:* R.N. Oddy *et al. Information Retrieval Research*, London: Butterworths, 172–191.

Paice, C.D. and Husk, G.D. (1987)

Towards the automatic recognition of anaphoric features in English Text: The impersonal pronoun "it". *Computer Speech and Language*, **2**, 109–132.

Paice, C.D., Black, W.J., Johnson, F.C. and Neal, A.P. (1993)

The construction of literature abstracts by computer. Final Report to the British Library R & D Division. University of Lancaster, Department of Computing.

Using text structure and text planning to guide text summarization

John A. Bateman

GMD–IPSI, Dolivostr. 15, D–64372 Federal Republic of Germany
(Also on extended leave from USC/ISI, Los Angeles)

1 Introduction: an application experiment

This paper will briefly outline an ongoing experiment being carried out at the GMD institute IPSI in Darmstadt, concentrating on the text generation aspects of the experiment and their possible relation to 'summarization'. The experiment combines work from three distinct areas: the 'editor's workbench' under development by the PAVE–group within IPSI as part of the European Community funded RACE–project EUROPUBLISHING, the text analysis component KONTEXT developed by the KONTEXT–group at IPSI, and the text generation component KOMET–PENMAN(ML) (KPML) under development by the KOMET group at IPSI with input from University of Sydney.

The experiment envisages the following scenario. An editor of a large–scale publication is gathering information from many source articles, considering how they are to be presented, what overview information can be given, etc. The editor's workbench supports this work by providing a graphically oriented object network editor, where objects can be displayed in a variety of styles. This workbench is implemented and has already been very favorably received by some potential end–users. The current application domain of the workbench is the 'Dictionary of Art': a large publication being prepared by MacMillan publishers. The functionality of this workbench is now being augmented. In particular, we include:

- deep analysis of incoming articles, producing a semantic representation of the content,
- presentation of that semantic representation in textual form.

These functionalities give rise to 'summarization' of various kinds, although summarization itself is not targetted as an independent task. This raises a number of issues concerning summarization: e.g., *is* summarization an independent task? Do 'summaries' have particular linguistics properties that need to be captured independently of other types of texts? Is there any difference between 'summarization' and, e.g., text generation as a whole? – since text generation is always of necessity selection of information to be expressed and cannot assume that information is anything but a 'summary' of the total information that could be expressed.

2 Some different kinds of summarization in the scenario

The places in our experiment where something similar to 'summarization' functionalities can be found are as follows.

The approach of the text analysis system is to provide deep semantic modelling of selected linguistic fields: e.g., the field of change–of–possession, motion, creation, etc. This deep semantic modelling is carried out on the basis of work such as [Kunze, 1991, Kunze, 1993], which now has received computational implementation [Firzlaff and Haenelt, 1992]. The construction of objects in the editor's workbench object network therefore proceeds by picking out of texts analyzeable sentences concerning one or more of the already handled semantic fields. In the present experiment, sentences concerning 'creation' have been targetted. During the inputs of source article texts, the constructed semantic network grows by addition of all facts concerned with creations of art objects, buildings, etc. We thus have a 'filtering' effect on the input, which can also be interpreted as internal summarization with respect to a selected–topic.

The editor's workbench itself supports graphical navigation according to the user's interaction with the object network. At any time, not all of the network is in view and certain types of relations (specified by or for the end–user) may be visible (view 'styles'). This offers 'summarization' of the local contents of the object network according to the user's immediate interest.

Finally, the editor's workbench may pass a request to the text generation component for the generation of a natural language textual expression of information in the object network. In addition to the constraints of what information is to be found in the object network, and the starting point of local interest fixed by the user during graphical navigation, the architecture of the text generation component also enforces a kind of summarization behavior since it seeks to strongly constrain the information that will be utilized in any text. This is to ensure that the text generation process is not overrun by the information to be expressed. The size of the object network containing information can be expected to grow explosively over the next year: the text generation process has, therefore, to bring to bear powerful constraints for restricting the information that it needs to access. This process probably comes closest to what is usually meant by summarization.

3 The text generation architecture

The current architecture of KPML attempts a full implementation of a systemically organized natural language architecture. In the spirit of, for example, [Cross, 1992], the systemic organization of [Halliday, 1978, Matthiessen, 1992, Martin, 1992] is used as the basis for all levels of linguistic information in the system including morphology, grammar, discourse semantics, register and genre. This computational architecture is under development with input from a number of cooperative research projects; these include on–

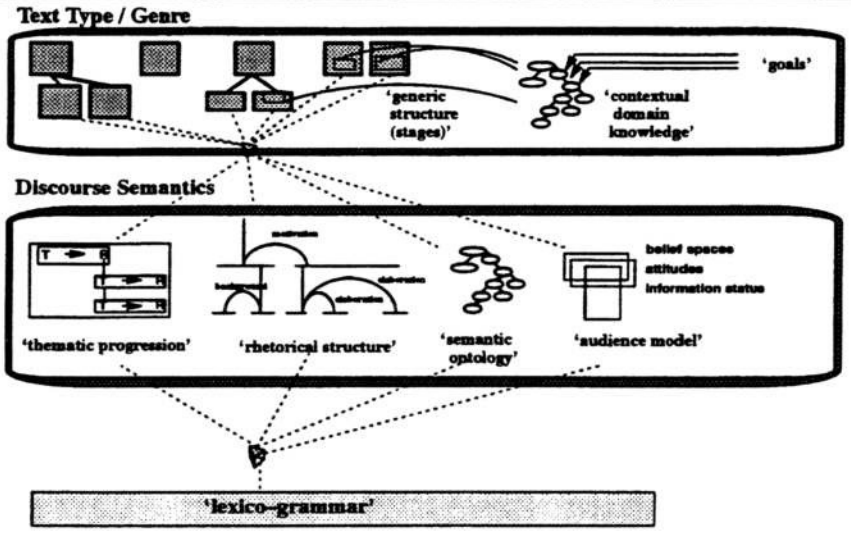

Figure 1: KOMET-PENMAN(ML) architecture

going work on 'register' (e.g., [Bateman and Paris, 1990]), work on multilinguality in text generation [Bateman *et al.*, 1991, Bateman *et al.*, 1993], and work in the EC–funded basic research project DANDELION (EP6665: Commission of European Communities).

The architecture represents a natural extension of work such as that of [Hovy *et al.*, 1992], moving in the same spirit towards a theoretically more homogenous treatment of textual phenomena.

In the present context, what is central concerning the architecture is its reliance on 'text type', or genre. Text generation only begins once a text type has been selected. This selection brings with it constraints both on the type of information to be selected from the knowledge base and its overall ordering into 'generic stages' to be presented over the text. This level of description corresponds exactly to the schemata originally used in [McKeown, 1985]. However, as suggested by the systemic orientation, schema are not unanalyzed wholes but are themselves the consequences of a classification hierarchy of a similar type to that deployed in the grammar. Classifying the genre gives rise to constraints on the generic structure that appears, just as classify within the grammar gives rise to constraints on the syntactic structure that appears. Each generic stage is then realized further by classification by the discourse semantics: a level of organization treated in depth in [Martin, 1992]. The result of this is a sequence of 'micro–semantic' specifications that can be passed to a systemic grammar: in our case, systemic grammars of German, English (the Nigel) grammar, and Dutch. The architecture is shown in Figure 1.

In our present work, we have been focusing on biographies of various kinds. Such texts have a relatively stable structure at the level of generic organization. Examples of texts recently generated are:

```
-- Behrens began his professional career as a painter. -- He attended
the art schools in Duesseldorf and Karlsruhe in 1886 - 1889. -- He studied
in Munich in 1890 with Kotschenreiter. -- He gave up painting after 1900.
-- He took up architecture in Darmstadt.
```

or

```
-- Behrens's principal activities were architecture and industrial
design. -- He made electrical appliances and flasks for mass production
in a glass works. -- Behrens started his career in Darmstadt in 1899
as an architect. -- He built the high tension plant and the turbine
factory for AEG in 1908 - 1910. -- He built a housing area for the
workers of AEG in Henningsdorf. -- He created a number of monumental
buildings and the German embassy in St. Petersburg.
```

with equivalents for German and Dutch.

Importantly, these texts were generated entirely on the basis of the object network in the editor's workbench knowledge base. The text planning is still very simple: for example, the granularity of facts in the knowledge base is largely taken over in the granularity of the events presented in the text. But this is in no way hardwired in the architecture; all such 'decisions' are consequences of the classification variations represented in systemic networks at the various levels of abstraction supported.

4 Summarization?

The question can then be raised, are the above texts and those like it 'summaries'? In a certain sense they clearly are: although they are not summaries of particular texts. They gather information from the knowledge base in order to fill out a text that has the structure and content of a subtype of biographies. To the extent that a biography can be said to be a summary, then these are summaries. But the same can probably be said regardless of the kind of text we generate. The same processes will be followed.

Is there then a separate kind of text type that we can label 'summary'?

Our work on biographies proceeds as follows. First, analyses of biography texts are undertaken to yield a 'register profile' of the text type. The function of biographies is also classified against a growing network of genre–level options. This needs also to be done for texts that would be described as summaries. However, whether there is additional information arising out of the fact of being a summaries that would not already be subsumed elsewhere is unclear. This can only be answered by empirical studies. If there is such an identifiable text type, then we can enter it into our genre level resources just as with any other text type. What is important is that text level constraints on the content and form of texts are an integral part of the architecture. As long as that is acheived, then text generation mechanisms would appear a natural candidate for constructing summaries.

References

[Bateman and Paris, 1990]

John A. Bateman and Cécile L. Paris. Constraining the development of lexicogrammatical resources during text generation: towards a computational instantiation of register theory. In Eija Ventola, editor, *Recent Systemic and Other Vietus on Language*. Benjamins, Amsterdam, 1990.

[Bateman *et al.*, 1991]

John A. Bateman, Christian M.I.M. Matthiessen, Keizo Nanri, and Licheng Zeng. The re–use of linguistic resources across language in multilingual generation components. In *Proceedings of the 1991 International Joint Conference on Artificial Intelligence*, Sydney, Australia, volume 2, pages 966 – 971. Morgan Kaufmann Publishers, 1991.

[Bateman *et al.*, 1993]

John A. Bateman, Liesbeth Degand, and Elke Teich. Multilingual textuality: Some experiences from multilingual text generation. In *Proceedings of the Fourth European Workshop on Natural Language Generation, Pisa, Italy, 28–30 April 1993*, pages 5 – 17, 1993. Also available as technical report from GMD/Institut für Integrierte Publikations– und Informationssysteme, Darmstadt, Germany.

[Cross, 1992]

Marilyn Cross. *Choice in text: a systemic approach to computer modelling of variant text production*. PhD thesis, School of English and Linguistics, Macquarie University, Sydney, Australia, 1992.

[Firzlaff and Haenelt, 1992]

Beate Firzlaff and Karin Haenelt. On the acquisition of conceptual definitions via textual modelling of meaning paraphrases. In *Proceedings of the fifteenth International Conference on Computational Linguistics (COLING–92)*, volume IV, pages 1209 – 1213. International Committe on Computational Linguistics, 1992.

[Halliday, 1978]

Michael A.K. Halliday. *Language as social semiotic*. Edward Arnold, London, 1978.

[Hovy *et al.*, 1992]

Eduard Hovy, Julia Lavid, Elisabeth Maier, Vibhu Mittal, and Cécile Paris . Employing knowledge resources in a new text planner architecture. In *Proceedings of the 6th International Workshop on Natural Language Generation, Trento, Italy, 1992*. Springer–Verlag.

[Kunze, 1991]

Jürgen Kunze. *Kasusrelationen und semantische Emphase*, volume XXXII of *Studia Grammatica*. Akademie Verlag, Berlin, 1991.

[Kunze, 1993]

Jürgen Kunze. *Sememstrukturen und Feldstrukturen*, volume XXXVI of *Studia Grammatica*. Akademie Verlag, Berlin, 1993.

[Martin, 1992]

James R. Martin. *English text: systems and structure*. John Benjamins, Amsterdam, 1992.

[Matthiessen, 1992]

Christian M.I.M. Matthiessen. *Lexicogrammatical cartography: English systems*. Technical report, University of Sydney, Linguistics Department, 1992. Ongoing expanding draft.

[McKeown, 1985]

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

Computational Resources II

Rapporteur: Elisabeth Maier

From Text Objects to Summaries

Ralph Weischedel

In the talk a system was described that automatically generates summaries out of multiple sources, i.e. potentially out of more than one document. The process of summarization is divided into the following subtasks:

(1) the understanding of documents and the object–oriented representation of their meaning (data extraction). The module fulfilling this task (PLUM) has so far been applied to 3 domains, 2 languages and a variety of knowledge sources (e.g. various wire services, etc.)
(2) the selection of a relevant information subset from the data base
(3) the automatic generation of a summary out of a subset of relevant information using a text generation system. he system used for this purpose – SPOKESMAN – has also been applied in many different environments (i.e., air traffic control, battle management, etc.).

Taking the two sample texts of Karen Sparck–Jones, the speaker showed how summaries have to look different for various reader types (e.g. for somebody being interested in the financial markets a summary of the pedagogical text would be NIL).

The speaker outlined some future research topics for automatic summarization as e.g.

- event/entity reference within documents
- data fusion across documents
- user customizability in all three summarization substeps
- summaries across documents (describing the same object)
- generation from fused data

The issues addressed in the discussion were:

- is knowledge representation THE bottleneck for the implementation of an automatic summarization system ?

  In the following discussion it was shown that knowledge representation (KR) as far as the

development of KR systems is concerned is not very problematic. The adequate representation of textual content, on the other hand, is still an unsolved problem which needs a lot of research. Knowledge acquisition was mentioned as another bottleneck for the implementation of summarization systems.

- inhowfar is a system as the one described transportable to other domains and other applications ?

  The proposed system has already proven to be transportable to other applications (see above). In general, it was felt to be more appropriate to develop a tool for specific purposes which could then be scaled up for the use in other frameworks than to build a general purpose tool which is only able to treat a small amount of problems due to its generality. By using some specific representations (like e.g. shallow templates) a higher generalizability of the implemented tools can be achieved.

- like in previous sessions the speaker was asked for his (contrastive) definition of abstract and summary.

  A SUMMARY was defined as a collection of relevant information extracted from the source document(s) while an ABSTRACT was specified as its (their) more generalized (i.e. more abstract) description.

In terms of the Questionnaire of the Workshop the talk contributed to answering the following questions:

A4: How general–purpose can summarising strategies be?

AS1: Are current techniques necessarily application–specific?
As indicated above, summarization techniques developed for a specific application can usually be scaled up.

UA1: How important is it to tailor summaries to the individual user?
It is very important, since readers always read summaries with a specific intention in mind which has to be considered when composing a summary.

CR2: Can present template–based processing be developed to find as well as fill templates?
As shown in the talk templates can be used for storing extracted information and later for the generation of summaries.

From Text to Objects to Summaries

Ralph M. Weischedel

BBN Systems and Technologies
70 Fawcett St.
Cambridge, MA 02138

In this paper, we discuss a hypothesized approach to automatic summarization:

- partial understanding of a text that produces an object–oriented representation,
- selection of a subset of (the object–oriented) data of interest, and
- text generation of a summary.

"Understanding," clearly an overloaded term, is used here to mean processing the input to produce a representation of the meaning of the text. Since full understanding comparable to human performance is very far away, only partial understanding is near the horizon. Finding instances of pre–specified types of objects (e.g., persons, corporations, other organizations, locations, and products) and pre–specified types of relations among them (e.g. products of a company, the officers of a company, the location of a company's headquarters, and subsidiaries of a company or set of companies) is feasible, and has been an active area of research in the U.S. (MUC–3, 1991; MUC–4, 1992). Such data is typically represented in an object–oriented framework. We will illustrate this aspect of the approach based on the linguistic components of our PLUM data extraction system and its performance in formal evaluations in four domains and two languages (Japanese and English).

Traditionally, a "summary" should represent the main points, therefore implying some selection among the data found is required. From our perspective, a non–traditional notion of "summary" is even more interesting, and more feasible. Suppose the user can specify individual criteria for information that should go in a summary, e.g., summaries of innovation in consumer electronics products, or summaries of joint ventures involving any U.S. company. Such a user–specific profile can be represented as a filter on the object–oriented data that is extracted, yielding an object–oriented representation. We will illustrate this process using the template generation component of the PLUM data extraction system

While the result of the selection process could itself be termed a summary, just as a table of data summarizes a set of facts very succinctly, it is easy to envision pipelining the objects to a text generation system, such as SPOKESMAN (Meteer, 1989), in order to produce a textual summary, or even a mixed mode summary, employing text, graphics, and tables. We will illustrate this process based on text generation technology.

User Adaption – needs and strategies

Rapporteur: Edward Cremmins

Raya Fidel:

User–Centered Text Analysis

Initially, it was pointed out that summaries and abstracts would be difficult to evaluate unless their purposes or uses are known. For example, if they are to be used in information retrieval they can be evaluated with regard to use of synonyms and other terminology. Bailey'ss evaluative criteria (e.g. depth/scope, accuracy, validity etc) were recommended for use also in evaluating indexes, abstracts, and summaries.

Physicians, as information users, find that structured abstracts for reports of clinical trials are helpful because they are more likely to ensure that the authors include sufficient relevant information.

The remainder of the talk was devoted to two questions:

1. How can we tailor summaries to individual users?
2. What do we need in order to do so?

Issues addressed in the discussion included

a) the difficulty of attempting to teach authors how to write;
b) the effects of structural abstracting on abstracting and the writing of the rest of the paper;
c) the tailoring of summaries to a particular audience, particularly using automatic techniques.

Tailoring of summaries for users, either interactively or "on the fly" was suggested.

The pros and cons of tailoring automatically generated simplified summaries to users' needs were touched on briefly at the close of the discussion.

Nick Belkin:

On the Relationship Between Discourse and User Intention

User adaption was examined within the context of the question: "How do we find out about what we need to know?"

The interaction of people with text was elaborated through a review of the partial results of a study of computer science students who wrote essays and filled out a questionaire about their interaction with document texts during the process. Although they were not directly related to summarizing, the students' activities could have aided them if they were asked to decide whether a given text was worth summarizing.

User adaption was next considered as it relates to the function of authors of texts; the intentions of potential users of summaries of texts; and the relationships, or dialogue, beweeen these intentions.

The writing (discourse structure) in a scientific paper and the plausible intentions of its authors were analyzed. The understanding of discourse structure and the relationships between the interaction of both users and authors could facilitate choosing texts for summarization that are consonant with the communicative and domain–oriented goals of both parties.

User-Centered Text Analysis

Raya Fidel

Graduate School of Library and Information Science
University of Washington
Seattle, WA 98195, USA

My presentation is aimed at raising questions for discussion, not at reporting about the results of research.

Two central questions are:

- Summarizing text; for what purpose?
- Intelligent communication; for what purpose?

Traditionally, summarizing text has included abstracting and/or indexing. What is the purpose of each? Generally, abstracts increase the efficiency of information gathering. More specifically, they:

- Help users decide whether or not to read a document
- Give orientation to the user who is planning to read a document
- Provide an overview to a user who needs to keep up to date
- Serve as a source of information by itself and cut reading time
- Make information published in foreign language available
- Make indexable material more visible

Indexing is performed for one purpose only: to facilitate retrieval.

In time, however, the distinction between the two is becoming less and less clear. Abstracts are used for retrieval, and they may even include controlled vocabulary. Index terms are used to indicate the contents of documents, as users may look at the list of index terms to determine what a document is about, and they may include terms in natural language.

From the user point of view, some attempts have been made to increase the usefulness of abstracts to users. For example, it is recommended that informative abstracts include four parts, and in an order that is useful to users:

objectives and scope; methodology; results; and conclusions.

A more specific example are structured abstracts for articles reporting original data from clinical investigations with human subjects. Here, a committee of users recommended the parts to be:

> objectives; design; setting; patients (or participants); interventions; main outcome measure(s); main results; and conclusions.

We know more, however, about the use of summarizing for the purpose of retrieval. Further, indexing has been the focus of retrieval studies. For this reason I would like to limit my comments to the area of indexing.

Indexing as summarizing is viewed as a two–stage process:

1. contents analysis that results in the selection of the concepts to represent the document, and
2. translation, that is, expressing the concepts selected in the index language used by the information system, or database.

This approach is called document–oriented indexing because indexing is derived only from the text. Thus, while the index language is supposed to reflect the user language, indexing by itself is confined to the text.

In real life, however, indexers rarely confine their analysis to the text; if they can, they consider the users and their information needs. This is not a new idea to information science. Brenner and Moores required a user–oriented indexing when they invented the edge–notched cards in 1958, and many others called for the need to consider user needs and requests. Soergel presented a complete approach and a method in 1985. In request–oriented indexing the indexer

1. gets first familiar with he document and its structure, and then
2. checks each descriptor in the index language and asks: would a person interested in (the descriptor's topic) be interested in seeing this document?

According to this approach, the index language is not only based on user language but it also "informs" the indexer about anticipated user needs. Moreover, the indexer is not limited to the document's text, and at times may need to use additional sources to determine whether or not a descriptor should be assigned to a document. This method is likely to improve retrieval performance, but it might be impractical because of the size of index languages. There are solutions, however: one may select a limited number of descriptors to create a checklist, or display the index language in a hierarchical structure, so indexers can avoid checking whole classes of descriptors that might not be pertinent to the document being indexed.

These are compromises, however. In addition, the request–oriented approach does not address situational changes, because indexing is a–priori. A better approach is to have a dynamic system that can index ad–hoc, tailoring its indexing to the specific or even momentary needs of a user or a request. This can be executed only with automated techniques.

In reality, however, most automated methods are developed with the aim to produce a *global* method; one that works for all users at all times. It might be useful to change direction and investigate how automated methods can perform indexing according to specific needs.

I do not have a proposal; rather, I would like to start a discussion. It is clear to me that before we even begin to attempt to change the tide, we must understand user needs better. Very little is known about end–user searching. Some examples of possible request characteristics, however, may start the discussion:

- Some requests may require documents dealing with a topic on a general level, others may require the specifics; can we provide this distinction in indexing?
- What to do if a topic is not explicitly mentioned in a document but is relevant?
- Some users are more knowledgeable about the topic than others.

I hope this presentation will generate a discussion about how methods of automated indexing can be sensitive to individual users and their needs.

On the Relationship Between Discourse Structure and User Intention

Nicholas J. Belkin

School of Communication, Information & Library Studies
Rutgers University
4 Huntington Street
New Brunswick, NJ 08903 USA
belkin@cs.rutgers.edu

ABSTRACT

The general issues being discussed at this seminar are:

- how can we best determine what aspects of a text should be included in a summary of that text; and,
- how should those aspects be optimally presented to the potential user of the summary?

In this presentation, I argue that a reasonable way to approach both of these problems lies in considering:

- intentions of authors of texts;
- intentions of potential users of summaries of the texts; and
- the relationships between these intentions, considered as a *dialogue*.

To address these three points, we have primary recourse to, respectively:

- concepts of text discourse structure;
- empirical investigations of types of, and goals for, uses of summaries; and,
- concepts of human interaction with text.

It is of use for us to have knowledge of the communicative intentions of the authors of texts, in order for us to understand what aspects of those texts the authors themselves considered important, significant, novel, or otherwise worthy of some reader's attention. Knowing this will give us at least some hints as to what aspects of the text are reasonable candidates for inclusion in any summary of that text, on the grounds that whatever the author of a text thought worthy of attention is also likely to be thought worthy of attention by some intended readers of that text. Knowing this will also be useful in deciding how to structure summaries, since it will suggest how best to relate the concepts

in order to achieve the authors' intended effects.

The communicative intentions of authors of texts are demonstrated, at least to some extent, in the structures of the texts which they produce, in terms of significant concepts, intended audience, and desired impact.

It is of use to have knowledge of the intentions and goals of the potential users of the summaries of texts, in order to understand to what kinds of purposes the summaries will be put. Knowing this will suggest, for instance, which concepts of the texts are likely to be relevant to the users in their decision–making or problem–solving activities; at what level of detail such concepts should be presented; and, how to structure the presentation of the selected concepts for maximally efficient and effective use in accomplishing the users' goals. Knowledge of users' intentions and goals can be gained from investigations of how people use the summaries which are available to them, and of the tasks and circumstances which lead people to seek texts, and summaries of them.

It is of use to have knowledge of the relationships of authors' intentions and users' intentions, in order to understand how precisely to choose the appropriate aspects of the texts, and to structure them, in ways that are consonant with the goals of both parties. In addition, such knowledge is necessary just in order to understand the intentions and goals of each party individually. For we really understand the intentions of authors only by virtue of the interpretations of their texts by readers, and we really understand the intentions of readers by virtue of how they interpret the texts, and in particular respond to the intentions of the texts, whether originals or summaries, with which they interact. It is thus in the interaction between reader and text that we find knowledge of the communicative relations between user intentions and author intentions.

For the purposes with which we are concerned, it seems reasonable to construe the interaction between user and text as a conversation or dialogue, for this then gives us a means to understand just how people go about establishing communication with authors through texts. In such a conversation, we can say that each party creates the role of the other, through a process of symbolic interaction, in which people interpret the meaning of texts not just in terms of their "information content" but in terms of what the author intended to communicate, and to whom the author is establishing some relationship. Through this process of interaction with text, it is possible for users to gauge the extent to which there is some intersubjective intentional alignment between themselves and the authors; which is how the meaning of a text is constructed. By studying this interpretative process, we have the means for gaining knowledge about how user and author intentions, and the relationships between them, could be used to construct and structure summaries of texts which are relevant to the communicative and domain–oriented goals of both participants.

I will attempt to demonstrate the potential utility of this approach to summarization by means of an extended example analysis of an author/text/ reader situation.

Evaluation Methods for Summarization

Rapporteur: Bill Black

Bruce Britton proposed that the first step in devising an evaluation strategy for summarising and abstracting systems or human practitioners alike should be to develop a classification of summaries according to requirements on the lines of that proposed by Hutchins. This breakdown uses purpose (informative, indicative or evaluative), information quantity (comprehensive or selective), and audience (domain specialist/general) as course dimensions.

Brigitte Endres–Niggemeyer proposed that the distinction between indicative and informative abstracts was redundant, proposing that the former were merely a pathological case of the latter. This did not meet widespread consensus.

Harold Borko, besides sticking up for the classical distinctions pointed out the distinctions in abstracting practice made necessary by subject domain.

Raya Fidel questioned the notion that a typology was a necessary preliminary to evaluation, which could consider such intrinsic factors as comprehensibility or readability, as well as extrinsic criteria such as fitness for purpose, and even conformance to editorial prescriptions. She also remarked on the proposed marginalising of indicative abstracts. Some objections were raised to the suggestion that conformance with editorial policy was a valid evaluation criterion, since it did not take purpose into account.

Paul Jones spoke against the elevation of informativeness as a criterion in itself, separate from considerations of fitness for purpose.

Mark Maybury emphasised the measurability of any identified criteria and the complexities of evaluating human against machine summarisation.

Brigitte Endres–Niggemeyer wanted to differentiate summaries and abstracts in purpose.

Udo Hahn expressed a wish to move towards formal criteria, such as distance measures in conceptual graphs. Metrics are needed.

Kathy McKeown observed that criteria for evaluating summaries differed from that for abstracts but there seemed to be a continuum between the two.

Karen Sparck–Jones recalled that Brice Britton had emphasised audience and purpose factors more than "input factors", and pointed out that using guidelines would be effectively a stand–in for evaluation based on what end–users want or need. ("Stand–in" was imperfectly heard and this gave rise to much mumbling).

Raya Fidel wished to emphasise distinctions in user populations.

Brigitte Endres–Niggemeyer quoted some data that abstractors often deviate from guidelines, often for good reasons, and also pointed to a neglect of the study of abstract use in relation to non–verbal communication.

Bruce Britton emphasised the value of categories as a divide and conquer method in evaluation experiment design.

Harold Borko, seeming impatient with the notion that abstracting journal editors were irresponsible in neglect of user needs, asserted that they receive regular feedback from subsribers and editorial policies do change.

Kathy McKeown then caused the discussion to change gear by making two concrete proposals for concrete evaluation protocols:

1. Take a corpus of source texts and summaries and compare the observable characteristics of human and system output along some distance metric.
2. A task model for evaluation in which a user is posed a problem involving the use of the summaries, and evaluation can be based on task completion and completion time.

Several speakers voiced objections to the procedure of comparing human and system abstracts for a variety of reasons not clearly articulated. Among these, Udo Hahn also emphasised the non–discrete nature of the variation in purpose to which abstracts and summaries are put.

Brigitte Endres–Niggemeyer reported on a recent meeting of a German society for social science information which had revealed an absence of abstract guidelines, and also proposed longitudinal studies of individual researchers's access to information.

Karen Sparck–Jones then proposed to classify this as a naturalistic (rather than investigator–set) variant of Kathy's Task–Based Model.

Rosemary Gläser introduced the problem of multilingualism of abstracts, and whether in the production of second language abstracts a transitory text is produced.

Liz Liddy epressed a strong preference for user needs study over a Type 1 study.

Raya Fidel sought to rescue the discussion from despair about defining evaluation methods by suggesting that there was a great deal of expertise in user studies that had revealed remarkable consistency in factors cited for user selection decisions. It is probably not necessary to devise experimental designs for evaluations.

Donia Scott warned about designing neat experimental methods without a prior clear understanding of criteria.

Mark Maybury pointed out that often the end users of information resorted to a variety of human intermediaries, both subject and information experts.

Karen Sparck–Jones drew lessons from IR practice, where the impact of automation had not been a direct substitute for human practices but had blown away conventional wisdom about indexing languages. She went on to propose naming Kathy's two protocols "extrinsic" and "intrinsic" evaluation respectively, and to characterise the problem of extrinsic evaluation as one of correlating contextual with textual factors.

Many contributers voiced agreement with the undesirability of direct human–machine comparisons, then Nick Belkin gave details of an empirical study that posed a problem involving access to literature and coded free–form answers against a set of 6 major facets of evaluation. (More detail is given in his Thursday afternoon talk).

Brigitte Endres–Niggemeyer pointed out that abstractors also have an obligation to authors of texts.

Ed Cremmins reported some background to the drafting and redrafting of the ANS on abstracting. Of 75–100 experts in the field, most had not thought extensive redrafting necessary, but a subset had been very vocal about extensive changes and a third revision was now under way, but no–one ever had a problem distinguishing the indicative from the informative abstract.

Nick Belkin asserted that goals lead to criteria, which in turn lead to measures – classic IR methodology.

Mark Maybury mentioned the need for measuring the improvements in the evolution of a system, which was more pressing for system developers than comparisons between systems.

Kathy McKeown reiterated a need for some framework of evaluation, however premature, despite

fears of a MUC–style perversion of research directions.

Raya Fidel claimed that qualitative analyses of system underperformance would be much more revealing than contest–type evaluation.

Karen Sparck–Jones pointed out that that activity was a necessary reflection on results, but not necessarily a substitute for hard facts, and in reaction to Donia Scott's reiterated warning about evaluation that was careless over criteria, said it was possible to be too perfectionist about evaluation.

Brigitte Endres–Niggemeyer was with Donia Scott and proposed a "start–up methodology" that would run qualitative pilot studies.

Harold Borko came back to the purpose to which abstracts were put – filtering, surrogate, orientation, and the value of introspection about these purposes when considering evaluation.

Hans Strohner observed that the variety of levels of understanding that an end user could seek or arrive at made the experimental set–up complex, and the simpler starting point would be to investigate the process of understanding an abstract.

Ed Cremmins could recall little in the literature on evaluation of abstract quality apart from readability studies, with a notable exception of Carol Tenopir of the University of Hawaii.

Rena Fidel made the final contribution to the discussion by observing that it is possible to find correlations between situation, task and requirements on the one hand an abstract characteristics on the other, once these relations have been validated with real users.

Open Problems in Text Summarizing:

Question List

We asked all participants to consider the following questions to identify important issues.

The answers can be found in the printed long version and in the electronic version, see note to the availability of the seminar's material.

A. GENERAL, for the workshop as a whole:

- **A1.** Can, and should, text summarising be decoupled from text interpretation?
- **A2.** Does summarising depend on recognising global text structure of a particular kind?
- **A3.** Can effective summaries be obtained with purely linguistic processing?
- **A4.** How general–purpose can summarising strategies be?

B. THEME–SPECIFIC:

HS – Human Summarising:

- **HS1.** How far does human summarising depend on task–specific training?
- **HS2.** What specific aspects of human summarising are best candidates for automation?

AS – Automatic Summarising:

- **AS1.** Are current techniques necessarily application–specific (for text type, subject domain, user need)?
- **AS2.** How well do current techniques capture large–scale text structure?

RD – Related Disciplines:

- **RD1.** How do human discourse processing strategies bear on summarising?
- **RD2.** What discourse properties are most important for summarising?

UA – User Adaptation:

- **UA1.** How important is it to tailor summaries to the individual user?
- **UA2.** What user features, and tailoring techniques, can we most usefully exploit?

CR – Computational Resources:

- **CR1.** How can current sentence processing methods help identify significant text content?
- **CR2.** Can present template–based processing be developed to find as well as fill templates?

Answers

Nick Belkin

A1.
No, summarization is of necessity, and in principle, a form of interpretation.

A2.
No, there are probably a variety of global and local text structures, the recognition of which support summarization. There appears to be no argument supporting the unique significance of any one of these to summarization.

A3.
Yes, if 'effective' is suitably defined, and if 'purely linguistic' means within the context of specific constraints. But, if effective means effective in general, and if purely linguistic means context free, then no.

A4.
It seems that summarizing strategies at the level of concept identification and structural characterization could be general to a variety of text types and intended users.

AS1.
They seem to be, in practice, if not in principle.

AS2.
Not too well.

RD1.
Explicitly, by suggesting which aspects of text are significant to persons in specific circumstances, by suggesting how these aspects can be identified, and by suggesting how they can be structured for effective presentation.

RD2.
Structure, in all its forms, and intention.

UA1.
Not at all. On the other hand, it is important to understand what aspects of texts, and of summaries of texts, are significant to individual users, in order to learn how to tailor summaries to specific kinds of uses, and for classes of users.

UA2.
The most significant user feature for text summarizing, is the goal or intention of the user which led the person to request a summary. Other relevant user features include such issues as domain knowledge, linguistic ability, and familiarity with specific text structures.

Bill Black

A1.
For the short term, it may help to decouple summarisation from interpretation, but there seem to be upper limits on extraction–based methods that don't interpret and then generate summaries.

A2.
Such structures appear accessible only in very limited text genres, such as experimental reports in some sciences, that follow a predictable pattern. Theories about macrostructures don't seem well–grounded in lower levels of analysis.

A3.
Partly a matter of definition. If lexical semantics is taken to be part of pure linguistics, then maybe, but the boundary between that and world knowledge is not very clear. However, it would seem necessary to have some encoding (however procedural or implicit) of the purpose of a summary and of conventions about its relation to the original text.

A4.
Well, low performance methods such as those based on extraction can work on a wide variety of texts. Knowledge–based methods will inevitably depend on some resources that have to be extended when the domain changes.

AS1.
The simpler and more heuristic the technique, the more it can be applied application independently. Linguistic and KB techniques will for some time be application dependent, if only because of the cost of the knowledge acquisition and the need to get some evaluation started before building full–scale systems.

AS2.
Large–scale text structure can only be captured reliably at present via mark–up done explicitly to signal structure by the author. Not the most academically interesting part of the problem. Linguistic methods are fairly limited to within–paragraph structure at the most.

RD1.
Well, the low–level processes have to be reproduced if texts are to be given reasonably interpretations by linguistic means. I'm sure observational studies must be very valuable in tackling higher–level discourse processing.

RD2.
Focus tracking is important in reference processing which is in turn an important element in low–level interpretation. Not so much to say about higher level structures.

UA1.
Really a matter of application requirements. If producing an abstracting periodial, not very.

UA2.
Well this is now quite an active topic in the generation community. I'm sure others have more to say.

CR1.
Not at all in themselves, but as a preliminary to either statistical or pattern–matching methods of identifiying significant text content, they should improve precision/recall, or at the least enable the construction of more modular/reusable systems.

CR2.
See CR1. Template matches should be based on semantic abstractions and not literal patterns.

Bruce Britton

A1.
I don't think text summarizing can be decoupled from text interpretation, although it would be much easier to do if it could.

A2.
Not only does summarizing depend on recognizing global text structure of a particular kind, but whether a summary is possible at all depends on which structure is recognized, because only some structures can be summarized at this time.

A3.
I don't see how effective summaries can be obtained with purely linguistic processing, because language can't be interpreted without world knowledge, world knowledge is outside "purely linguistic processing", and text interpretation is needed for text summarizing.

A4.
The summarizing strategies I know of are relatively special purpose, in that they can only work on quite specific text structures, e.g., causal and procedural texts.

RD1.
The human discourse processing strategies which are most important for my paper are those needed for constructing a situation model using materials from the text and the reader's prior knowledge. This covers a pretty wide range of strategies.

RD2.
The discourse properties most important for summarizing for my paper are the causal and procedural models that gave rise to, and are expressed in the text.

Ines Busch–Lauer

A1.
No, text summarization should not be decoupled from text interpretation. All reading, writing and speaking as well as the summarizing process always include the individual reflection of language and of the reference text in the human mind and are thus an interpretation which is determined by social, cultural, cognitive and linguistic conventions. Communicative function, purpose, addressee and subject area do influence the summarizing process.

A2.
Linguistic studies on various types of summaries reveal that the summarizing process often relies on the global text structure of a particular kind of text (genre, text type). This can be observed in abstracts of experimental research via recurrent "moves".

A3.
Effective summaries cannot exclusively be obtained with purely linguistic processing. Summaries are a result of purposeful text interpretation which relies on individual or cross–individual world knowledge, intuitive knowledge about texts and their structure.

A4.
I think that there exist some fairly general summarizing strategies, e.g. mark topic sentences. But global strategies must be specified according to the purpose of summarizing, addressee, subject area, genre of reference text, type of research performed (empirical, experimental, etc.).

HS1.
There is a close relationship between task–specific training and the process of summarizing. Trained experts both in the area of summarizing and the subject area will perform much better than inexperienced students. Teaching experience shows that summarizing as a process should be dealt with in the native and the foreign language education.

HS2.
Possibly

- thematic text progression according to the macroproposition(s) of the reference/source text
- recurrent "moves" like introduction, purpose of text, subject matter, conclusion.

RD2.

Very important discourse properties for summarizing are: global text structure (macrostructure), visual lay–out, metadiscourse (metacommunicative elements signalling moves and steps in the written text), macropropositions and their thematic progression.

Fabio Ciravegna

A1.
I consider text summarising and text interpretation as complementary parts of the global task of the information processing. Those parts cannot be decoupled, for the second depends on the first exactly as the first depends on the second. As a matter of fact to correctly summarise a text one should first extract the correct information from the text; at the same time to extract the correct information one must know what it is going to be produced.

A2.
Yes: summarising depends on the recognition of all the text levels: from the physical format, to the linguistic, pragmatic and intentional levels.

Physical format
The recognition of the main structures of the text, such as the title (or headline), the eventual by–line, the abstract and the conclusion is fundamental. If not correctly recognised, the structure of the text may seem schizophrenic or clashing, for headlines, by–lines, abstracts and conclusions report information that is already present in the body of the text.

Linguistic and pragmatic structure
The history of the discipline shows that the recognition of the linguistic and pragmatic structure of the text is necessary to achieve high precision and recall rates.

Intentional level
The intentional level is the more difficult aspect to catch: in some ways it can't be taken into account in all details, but an effort in that direction is desirable, mainly for automatic summarising.

A3.
No; if summarisation is meant to be part of the global project of information processing I referred to in point A1, it can't be limited to pure linguistic processing, provided that in the linguistic processing you don't include the recognition of the user intentions, of the structure of the text etc. (see point A2)

A4.
In automatic summarising all the strategies should be general–purpose at the maximum extent. Unfortunately it seems to me that the current state of the art techniques don't allow such a possibility. What is desirable for the future is the definition of approaches favouring generality and powerfulness to allow:

- in depht analysis and production, when possible;
- a shallow approach otherwise.

The availability of suitable knowledge bases, dictionaries, grammars, etc. are important elements in that direction.

AS1.
The current state of the art approaches in information extraction show that the key point is to mix general purpose modules and application specific modules to maximise recall and precision (see for example Rau and Jacobs' GE NLToolset). Anyway I think we must separate the long–term research from the short–term research: some experiences (as the TACITUS project by Hobbs et al.) show that it is possible to achieve some interesting results using more general approaches that will give better results in the long term period; other experiences (as the FASTUS project again by Hobbs et al) show that more ad hoc techniques are suitable to get short term results (in the latter case the long–term results will come from the refinements of that ad hoc techniques).

AS2.
Not very well, unfortunately.

RD1.
RD2.
I think that one of the main human discourse processing strategies bearing on summarising is related to the discourse focusing, i.e. the ability of organising the information in the text in the correct way allowing:

- to understand the relevance of each information at every step;
- to resolve correctly the references among different parts of the text.

UA1.
As summarising can be viewed as part of the information processing, tailoring to the individual user is as necessary as it is in each human computer interaction.

UA2.
I think that the most important feature from an applicative point of view is the interest of the user, to allow him/her to receive correctly tailored summaries of the information he/she is interested in; otherwise the user must read through boring and not focused text wasting time and lowering attention.

CR1.
Current general methods use simple but powerful lexico–semantic pattern matching (see for example Jacobs' work at COLING90 or at the MUC conferences) that allow high precision and efficiency. Other interesting approaches, although limited to a specific application, use more powerful techniques based on semantic features based on some partial linguistic analysis (see for example Ciravegna's work at COLING92).

CR2.
No, I think that template based processing can't be developed to fill templates. They generally lack in precision and in capturing the user intentions. Although the presence of some good applicative results as the ones of the Hobbs' FASTUS project, I think they are not going to be the answer to the need of information extraction in the long–term period; it seems to me that they are a good answer for short–term results in that field.

Edward Cremmins

HS1
HS2
"Human Summarizing" is the only category of the questions in which I have expertise. The first question in this category (HS1) is well formulated, but I don't have any empirical data to offer to assist in answering it. However, I will keep it in mind, as I continue to prepare my talk for the seminar. Question HS2 is on automated summarizing with which I am only familiar indirectly through my reading.

Brigitte Endres–Niggemeyer

A1.
As far as I see, we always summarize from a standpoint. Therefore summarizing depends on a particular interpretation of a text. In addition, a specific summary–bound text interpretation strategy exists and is interesting: It avoids most text interpretation by concentrating on the passages that are really needed for the intended summary.

A2.
Yes, empirical data from professional abstracting shows that normal abstracting relies heavily on global text structures. It is harder to summarize unstructured or scarcely structured documents, although it is quite possible. Where global text structures provide no guidance, text macrostructures and domain knowledge are used.

A3.
No. No summarizing without knowledge processing.

A4.
There are fairly general summarizing strategies, e.g., "Pick topic sentences". Others are domain specific, e.g., in chemistry ("Keep chemicals") or history ("The last state of a development is its summary").

HS1.
Heavily. Experts perform much better than untrained students, among other things because they know better task–specific methods.

HS2.
My favourite candidates for implementation:

- From an empirical point of view, features of interindividual stability
- Frequently used core strategies
- Professional task–oriented text views
- Selective information acquisition strategies
- The human process organization
- Frequent document structures
- Core domain knowledge structures

AS1.

Not only the current ones, I fear. Human experts are "application–specific". They combine techniques that rely on specific features, e.g., text type, domain knowledge structures, or user interest. Nobody masters all of them. Systems that perform where humans are overwhelmed are far away.

AS2.

To my knowledge, no current automatic summarization technique deals really with large–scale texts, e.g., monographs or long reports, unless one includes good old sentence extraction methods.

RD1.

Summarizing is integrated into normal human discourse processing. Because they do most of the job, the most interesting human summarization strategies are:

- large–scale strategies of dynamic and selective text understanding
- schematic strategies working with text superstructures, domain knowledge, and personal profiles of recipients and authors
- semantic cutting–and–pasting strategies for target discourse production.

RD2.

Good text design: superstructure, visual layout, macrostructure, knowledge substance.

UA1.

User adaptation is important since summarization has to respect what the user wants to know and what she is able to absorb. Otherwise, the result will not be useful. However, it may not be necessary to tailor to an individual user all the time, because often, group profiles can cater for features of individuals, e.g., for persons in a hurry, or for computer scientists.

UA2.

Combined information seeking profiles of individual users seem most helpful:

- information absorption capacity?
- missing knowledge?
- existing individual knowledge system?
- target of information seeking / current task?
- learning strategies?
- ...?

Raya Fidel

A1.
Any answer to this question probably has epistemological origins. To me, any summarizing is an interpretation, even if it is performed with automated techniques. The final product of summarization is determined by the process, and the process provides the interpretation. It is highly unlikely that two different processes would generate the same summarization. Moreover, we should not worry about interpreting text; any text that is read by a human being is being interpreted. Thus, summaries (whether or not they include interpretations) are subjected to further interpretations when read by users. It might be useful to be able to designate the nature of the interpretation. Even more so if the interpretation is made for the potential users; for example, if an article in economics is summarized for the lay person, or one about the use of isotopes in geology for geologists.

A4.
Experience shows that the more general the technique, the less powerful it is. Subject domains seem to be an important factor. To me the question is: What parts in our strategies should be general, and what should be domain specific. For example, we may find that METHODS or PRINCIPLES for generating strategies can be developed and be applicable to all domains, but the strategies resulting from these methods and principles are domain–specific.

UA1.
It is difficult to answer this question because we know very little about how users employ summaries. But we do know that in retrieval systems, users (whether end–users or professional searchers) work very hard to adapt their search strategies and improve retrieval. This indicates that the more we tailor summaries to the individual user, the less the user would have to adapt to the system. It is difficult to determine whether it is important. Some may claim: as long as users adapt so well (an assertion yet unsubstantiated), why do we need to make this special effort and tailor to individual needs? On the other hand, if certain relevant aspects are lost in a summary, they will not be available to users, no matter how adaptive they are. It is important, therefore, to first find out how users employ summaries, and then tailor the text to individual needs as much as possible.

UA2.
This question can be answered only by research. For a long time information science has been speculating about users and their needs. Today, a growing number of users have direct access to information systems. Therefore, the feasibility of direct investigation of their seeking and searching behavior through direct observation has much increased. Such studies would point to user features that could be most usefully exploited, and to the most suitable tailoring techniques.

Rosemarie Gläser

A1.
In my view, text summarizing presupposes a thorough understanding of the text, which, to a certain extent, will result in text interpretation. Thus, text summarizing cannot be completely decoupled from text interpretation, unless one has a very superficial approach to summarizing.

A2.
Summarizing depends on recognizing the global structure of the text under analysis. The whole macrostructure and its individual paragraphs should be analyzed closely before summarizing.

A3.
Effective summaries will depend on profound subject knowledge and thorough mastery of linguistic techniques on the part of the processing person.

A4.
General–purpose summarizing techniques include rhetorical techniques, such as time order, space order, cause and effect, comparison, contrast, analogy, exemplification and illustration, and rhetorical functions, such as definition, classification, and description (cf. Louis Trimble et al. 1978, Trimble 1985).

HS1.
Human summarizing is acquired by experience; task–specific training is indispensably necessary.

AS1.
I am not familiar with "current techniques" in automatic summarizing but I suppose that they will vary according to the genre (in terms of John Swales) and to the text type (in terms of Egon Werlich: description, narration, exposition, argumentation, instruction).

AS2.
The term "large–scale text structure" calls for clarification. Do you mean 'superstructure'? Generally speaking, each "large–scale text" should be carefully segmented into structural units (chapters, chunks, conceptual and physical paragraphs) before summarizing.

RD1.
Establishing coherence in the process of listening comprehension and reading comprehension is an important prerequisite in human discourse processing and has a bearing on summarizing.

RD2.
A clear and logical elaboration of the subject matter in the source text, lucidity of its macrostructure, subheading and clear conceptual paragraphs are discourse properties which are most important for summarizing.

UA1.
Summaries should be multi–purpose text reductions and, in the first instance, should not be tailored to the individual user. In the last instance, the type of summary will depend on the material where it is published (abstract journal).

CR1.
CR2.
I am not prepared to answer these questions because I do not work in this area.

Udo Hahn

A1.
Yes!
From the computational point of view it seems reasonable to separate parsing (text interpretation) from *any* further transformational process (such as summarisation) since the text interpretation structures (I assume these are knowledge representation structures, no phrase markers, grammatical feature sets, etc.) can then be used as well for alternative transformational processes, such as translation, message routing, text knowledge extraction, etc.

A2.
Yes!
Any adequate account of text parsing depends on the proper recognition of local and global text phenomena (text cohesion and text coherence). As I will argue in my paper, lacking recognition of text cohesion phenomena will cause *invalid*, lacking recognition of text coherence phenomena will cause *understructured* text knowledge bases (TKBs). As TKBs representing the content of the text are the major knowledge source available for the summarisation procedure, it should be obvious that understructured TKBs produce poorer results than fully elaborated ones. Global text structure descriptions, e.g., properly determine the major topic(s) of a text, the major lines of topical development (topical thread), i.e. information that is vital for any successful summarisation attempt.

A3.
No!
Of course, the answer depends on what one considers as "purely linguistic processing", but any reasonable interpretation of this phrase leads me to a negative answer. The main reason why "purely linguistic processing" is not sufficient is that it lacks the provision of common–sense and domain–specific background knowledge.

A4.
I doubt they can be general–purpose at a global level, since many other system parameters (user interest, previous knowledge of the user, news contained in text, etc.) have to be incorporated into the basic information reduction machinery. However, if there is a "knowledge condensation kernel" we have some proposals to make with respect to those criteria that should be taken into consideration for the general–purpose part of summarisation procedures (see my answer to AS1).

AS1.
This is just a claim. But since our knowledge abstraction procedures only refer to fairly general criteria (terminological hierarchies, relevance weights, graph connectivity) I may defend this statement based on plausibility considerations. But I have no proof.

AS2.
Poorly, unfortunately!
Beyond the level of focus management, for which we have some working principles and systems, there isn't much available. Who knows of operational systems that parse into story grammars? into RST (some generators yet exist)? into HOBBS–style coherence relations? into KINTSCH/vanDIJK–style micro–macro propositions (excluding some work of Correira and Simmons)? There is a shimmer of light at the end of the tunnel when one considers the plot unit systems (Loiselle & Lehnert) or work on TAUs/TOPs/MOPs as proposed by various Yale people, but each of these proposals is fairly restricted on its own.

Jerry R. Hobbs

A1.
If they can be decomposed, there is no reason they should not be. However, it is unlikely that we will ever be able to produce a reasonable summary of a text without understanding it first.

A2.
The most promising approach involves the recognition of global text structure, since it is surely necessary to discover which text segments are dominant and which subordinate, to recognize the implicit generalizations that parallel segments instantiate, and to determine a coarser–grained description of a coherent sequence of events, among other things.

A3.
What do you mean by "purely linguistic processing"? World knowledge is certainly needed in order to interpret a text, and text interpretation is probably required for decent summaries.

A4.
Certain restricted genres of text may lend themselves to special purpose summarizing strategies, but the vast bulk of texts will probably require more general–purpose methods involving text interpretation.

HS1.
No opinion. I will be interested in learning at the workshop something about the economic functions of human summarizing.

HS2.
No opinion. I hope that some ideas about this question will emerge from the discussions at the workshop on how human summarizing is used currently.

AS1.
Not really.
There are at least three families of current techniques in use today:

1. Finite–state methods of text processing, recognizing phrases on the basis of linguistic information and piecing together patterns of phrases that are of interest for the task. The first step is application–independent; the second involves application–dependent patterns, but the mechanism for recognizing the patterns is application–independent and the patterns are easy to write for any particular application. However, while this technology is moderately good for information extraction, it is probably inadequate for text summarization.
2. Parsing, plus compositional semantics, plus type checking (i.e., computational linguistics as usual). The type hierarchy is domain–dependent, but the other processes are domain–independent. These methods probably do no better than finite–state methods in any application. The purpose of parsing is to recognize predicate–argument relations, and the purpose of recognizing predicate–argument relations is to make inference possible. But with these techniques, no serious inference is done. So most of the processing is wasted.
3. Abductive interpretation using a large knowledge base of world knowledge. The knowledge base is in large measure domain–dependent, but the inference techniques are domain–independent. While theoretically promising, this approach requires the substantial infrastructure of a knowledge base, and serious attention must be paid to developing efficient inference processes. These methods have been implemented only in small–scale research systems.

AS2.
Finite–state methods and computational linguistics as usual: Not at all. When "discourse processing" is claimed for such systems, what is meant is a few simple kinds of coreference resolution.

Abductive methods: Promising. But implementations are nonexistent.

RD1.
Probably too little is known about either human or computer discourse processing strategies for the distinction between the two to be useful. Any good psychological account of discourse processing has to be cashed out in computational terms, and any good computational account can be described more abstractly in psychological terms.

RD2.
Probably global structure.

UA1.
No opinion. This is an issue of how summaries are used, and whether broad classes of users exist.

UA2.
No opinion.

CR1.
Finite–state methods can identify certain patterns of interest, and it may be that the repetition of such patterns in a text indicates greater significance. But novel significant text content is beyond current methods.

CR2.
One can imagine techniques for specializing already existing, abstract template types as the templates are filled for a corpus of texts.

Paul Jones

A2.
I'm not sure that summarising does 'depend' on recognising global text structures but it only seems sensible to use all possible clues/pointers to important material whether using an automatic or manual summarising system.

A4.
I believe that there is a need here to make a distinction between strategies and systems. A particular strategy may be applicable to a number of domains/types of papers, however a system using the startegy may need doamin/text type details in order to perform the summarisation effectively.

It seems unlikely that a truly general purpose strategy can be developed. I would not expect the most effective strategy for summarising a paper on a chemistry experiment to be the same as the most effective strategy for summarising a philosophy text book. However the 'chemistry paper strategy' may be effective on papers which similarly report on empirical work, maybe from the fields of physics, agriculture or engineering.

AS1.
Some are, some are not. Should one of our present aims be to produce a totally generic solutio ? Would it not be more sensible to produce a number of satisfactory application–specific techniquies and move from these to a more generic platform? After all human abstractors specialise in particular domains, why shouldn't automatic systems specialise.

AS2.
I am not aware of any that do this particularly well.

UA.
Rather than answer the set questions I would like to ask another question. *"How are summaries used ?"*

I have anecdotal evidence from colleagues that agrees with my experience, namely that summaries have two principle function : an indicative function and a reminder function.

The indicative function is used on first aquiring the summary to decide if the full paper is worth reading/aquiring. If the document is read then the summary acts as a reminder as to its contents when the user refers back to it. If this is how summaries are normally used it seems that a lot of summaries are overburdened with details of results etc.

However a number of medical journals have moved over to using structured abstracts which seem to increase the amount of data provided and anecdotal evidence from an abstracting company I have dealt with suggests that their abstracts are often used as a major source of information.

If different application areas require differing levels of complexity/detail in the summaries they use and indeed if they are used in different ways, then it is necessary to tailor summarising techniquies/strategies dependant on what the user group requires. Perhaps the first stage of any work on summarising should begin with the question "How are these summaries going to be used?". I will be very interested to speak to anybody who has access to, or has seen any empirical evidence on the use of summaries.

Elizabeth D. Liddy

A2.
This is an empirical question and the answer is likely to be dependent on the particular approach to summarization which is taken. However, my experiences with recognizing global structure in documents leads me to believe that text structure is useful in at least some implementations. Perhaps more so in some text types than others. We have found with our work with newspaper texts, that the system's ability to differentiate between background and foreground information on the basis of structure can be useful in creating news summaries. Since the discourse model of newspaper texts is one in which the main story is told in non–contiguous units, the ability to recognize which text units comprise the main story to be summarized via the delineation of the text structure is essential. However, I have seen other work on different text–types which can differentiate between main topic and sub–topic without any explicit inclusion of structure.

Some of the early work in automatic abstracting done in information science depended quite explicitly on orthographically recognizeable indications of text structure. That is, for scientific papers, the first sentence from sections which were thought to be particularly indicative of content needed in an abstract were extracted for concatenation as a summary. The simple fact that humans frequently peruse the first sentence of orthographic sections in a text in order to get a sense of what's include, suggests the possible utility of the use of explicit structure.

A4.
At this moment in the evolution of automatic summarization, I think general purpose strategies are not really possible (but perhaps I will be pleasantly surprised by new work at the Dagstuhl). But again, this may depend on what is meant by summarization. That is, does a good summary, like a good humanly constructed abstract, devote proportionately the same amount of coverage to topics that is found in a full text, or does it summarize the main points only?

HS2.
The specific aspect of human summarizing which would lend itself most easily to automation is determination of the topics most talked about in the text. Although this might not be as easy as thought due to the use of anaphora can result in the surface level under–representation of main concepts. Additionally, synonymous references to a single entity, event or concept are easy for humans to collapse into a single topic, but require additional semantic processing by a system.

AS1.
Again, I am waiting to be pleasantly surprised by new work at the Dagstuhl, but my own experience with processing of various text types suggests that different types would require (or be

improved by) application–specific processing. At some level, the techniques are the same, but the particulars would vary. Again, a question of definition here, what is meant by techniques. Is this to be taken as 'approaches' ?

AS2.
I would like to attempt to respond to this question, but I do not know what the meaning of 'large–scale' is in this situation.

Elisabeth Maier

A1.
It cannot be decoupled for the following reasons:

- constructing a summary depends significantly on the comprehension of the source text. I.e. the problem of synthesizing the summarized target text includes an interpretation and analysis of the source text.
- text summarizing includes phases of revision, i.e. preliminary summarization products are continuously monitored, re–read and improved; revision includes both interpretation and generation.

A2.
Certainly ! Global text structures often give a clear indication where relevant pieces of text can be found. For example, conclusion chapters contain the core of all the arguments given in a scientific text. So, scanning such a chapter can in many cases be much more efficient than reading the entire text. The recognition of the text type, the recognition of its conventionalized text structure and the knowledge about the contribution of the various text units to the meaning of the text as a whole are essential prerequisites for efficient summarization.

A3.
I am not sure I understand this question. All operations on a text which are executed in order to construct a summary are linguistic. Some of these operations might be motivated or triggered by pragmatic or situational features but this doesn't change the linguistic nature of the single operations. Some summarizing operations like e.g. marking or exploitation of text layout can also be considered linguistic or can at least be seen as part of an extended discourse model which integrates textual with graphical presentation and interaction modes. So, to conclude, summarization can be considered an entirely linguistic process assuming that

(1) the notion of "linguistics" is fairly broad including also non–textual communication means and that
(2) linguistic processes can be based on pragmatic / situational grounds.

A4.
This question can be interpreted in various different ways:

- A4a. do summarising strategies depend on the text type ?
- A4b. do summarising strategies depend on the individual abstractor ?
- A4c. can abstracting strategies also be employed by other natural language processing devices ?

The first two questions can be answered using the same argument: it is very likely that for the various text types and the various abstractors a core set of summarisation operations exists. For individual genres and abstractors a set of specific operations can be added. Concerning the third question I would say that once a set of core summarisation operations is determined they can be incorporated in other systems; for example, some construction operators which are employed for the construction of a coherent summary out of previously identified relevant text units can be re–used for text generation purposes. Similarly, scanning procedures could be re–used in text understanding systems.

HS1.
Human summarising is not my field so I feel unable to judge.

HS2.
First of all those aspects and partial processes which have already been developed for other NLP systems and which therefore can be possibly re–used, as for example text comprehension and text generation tools. Besides that, I would say, that the automation of human summarizing steps depends highly on the advancement of research in exactly that field and on the granularity and formality of the models developed there. Since I am not an expert in the field of human summarisation, I am not able to judge the advancement in some subfields of summarisation with respect to a possible automation and implementation.

AS1.
As mentioned in the answer for the questions A4a. and A4b. above certain techniques can be expected application–specific while others are most likely common to most possible applications.

AS2.
First of all, I am not quite sure what is meant by "large–scale text structure". I assume that it relates to structures of very big texts, e.g. books. If we take macrostructures and relational approaches (e.g. RST) as sample theories for describing text structures we find that they also can be applied to describe large–scale text structures. Therefore, I think that there is no problem in describing large–scale text structures using current techniques.

RD1.
Among the human discourse processing strategies which come to my mind first is the ability to order relevant information in the most efficient way. Such an ordering implies that information is placed exactly at the positions where they are expected by the reader. So, for example, in a

scientific article we assume that the gist of the text can be found in the conclusion chapter while an introduction to or a circumscription of the problem dealt with is most likely to be found in a chapter titled "introduction" or "problem". For summarisation, such strategies of text production are obviously to be exploited. I assume that when a text is to be summarized which does not follow common text structuring conventions, information relevant for the summary is much more difficult to retrieve.

Similarly, the construction of the final summary out of text pieces which have been marked as relevant also uses some standard techniques of text production, as e.g. the production of cohesive links using referring expressions, cue words, lexical chains, etc.

In general, I think, that a summarising model which makes only scarce use of human processing strategies is much more inefficient than a model which employs such strategies.

RD2.
From my personal point of view, and this also due to my background, I guess, the most important discourse property is discourse structure. By discourse structure I mean both the sectioning of a text into units and the relationships holding between these units. Concerning the sectioning of a text we have to distinguish between semantically motivated units (i.e. text units which address one subordinate theme or topic and which can be considered a "closed" entity) and units which can be distinguished on the level of text layout (paragraphs, sections, chapters, etc., but also itemized lists, tables, figures. etc.).

UA1.
It is as important as tailoring **any** text to the potential reader, i.e. very important.

UA2.
Concerning user features and tailoring techniques (especially the latter) I am only aware of work done in the field of natural language generation (generation of descriptive texts, explanation generation, generation of tutorial dialogues [Paris, Suthers, Moore,..]); the techniques employed there, which influence both content and structure of the texts to be produced could well be employed for the synthesis part of a summarisation model. Still, more techniques have to be developed which handle user–specific aspects in text analysis and in recognizing information relevant for a summary.

Among the user features which have to be taken into account are (to mention only some)

- the knowledge of the user this feature influences the information to be provided in order for the abstract to be understandable; it also influences lexical selection as far as the use of a

specialized vocabulary is concerned.

- the interests of the user the interest profile of the reader influences the depth of elaboration of a summary; if the summary concerns a topic which is of high interest for the user it might be longer and perhaps provides a higher degree of detail than an abstract which is only of peripheral interest to the user.

CR1.
Here I suggest that current techniques for information extraction are considered ([Rau, Jacobs, Pattabirhaman, Ciravegna]).

CR2.
I do not understand the relevance of this question for the topic of summarization. Nor do I understand what is meant by "template–based processing" in that context. But I cannot imagine why one shouldn't be able to develop such a tool.

Mark Maybury

A1.
As interpretation is increasingly decoupled from summarization, an increasingly richer representation (e.g., to capture discourse structure, intentions) is required. Thus, an advantage of retaining coupling is the complexity of representation may be reduced. Another is that by having access to the original source this maintains linguistic constraints that can be used in summarization (e.g., location of information in text, lexical & structural choices).

A2.
Probably text type dependent (e.g., lead sentences found in newspapers suggest the importance of location over structure).

A3.
vice processing that includes "world–knowledge"? Determining saliency can depends on interests of user, what is know to be important in the domain, and what is important in the extra–linguistic context (e.g., current events).

A4.
Our experience with simulation summarization is that the strategies for ordering/grouping information are general (e.g., based on time, space, topic) but that strategies for selecting information can be both general (e.g., select the most infrequently occuring events) and domain–dependent (e.g., what is deemed an important event in this domain).

HS1
Helmut Felix Friedrich's comparison of macro–rules and structural approaches to human summarization suggests that domain–specific knowledge interacts with domain–independent "methods" of summarization.

HS2.
Those that are based on explicit (lexical, syntactic, semantic, structural) cues in the source text that are within the current state of the practice in computational linguistics. Of course, in genres such as journalistic writing, more sophisticated techniques might be outperformed by simpler techniques (e.g., taking the first X words), as a consquence of the nature of the genre (present important information first).

AS1.
yes.

AS2.
not that well. While it is easy to take advantage of explicit format (e.g., section headings), large scale text structure (rhetorical or intentional) is beyond the ability of most if not all current systems.

RD1.
The highly distributed, parallel, and iterative nature of human summarizing provides an interesting contrast to the, primarily, sequential processing in current systems.

RD2.
given/new distinctions, focus information, discourse/dialogue structure, intentional structure

UA1.
Important if there are multiple perspectives on a subject, or clearly distinguishable reader classes or individuals.

UA2.
Many, A difficult problem, however, will be deciding what to elide from the resulting abstract/summary as unwanted implicature or inferences will be necessarily user and context specific.

CR1.
statistical analyses (e.g., word frequencies), cue words, syntactic analysis, predicate–argument processing

CR2.
We need to computationally formalize what is meant by relevant and significant information.

Kathy McKeown

A1.
It can be, if one thinks of

- 1. Identifying in/extracting from the text just the information that one is interested in including in the summary, in place of interpreting the entire text.
- 2. Summarizing input data is another way that summarization can be decoupled from interpretation. In this case, generation of a summary can be done separately.

A2.
I don't think so. It can help, I'm sure, in identifying the information that is more important in the text but I don't think it's absolutely necesssary, particularly for certain tasks such as information extraction.

A3.
I doubt it. Certainly some identification of the important facts in the domain is needed and this requires domain knowledge.

A4.
Right now, it seems hard to develop a summarizing strategy that would work in any domain. What is done often seems to depend on the conventions of the domain.

AS1.
Yes. Current techniques usually expect a certain text type and make assumptions about the kind of information found in that type.

AS2.
Not well ...

UA1.
I think it is important to tailor the summary to the goal of the user. Can be a broad range of users with similar goals (e.g., seeking an article). Summaries definitely must be different for different goals.

CR1.
CR2.
I think an appropriate question that was not asked is: how well do current generation methods help in producing a summary? Most people ignore the generation side, but if it's text to text there's a lot more than just identifying the key points. Usually additional information is integrated into the summary as it fits. Current generation technology is ready in many aspects, but the need for complex sentences is not easily handled by many systems.

Sumiko Mushakoji

A1.
No. Even in automatic summarizing, that is purely processed by non–human, how can the products be recognized as "summarized" "texts" without people's text interpretation.

A2.
If that is a case of summarizing by the person other than the producer of the original text, they may use the surface structure of the original text (e.g. style, headings). Some empirical evidences in the cognitive (psychological) studies have demonstrated that, and in my sociological standpoint, the person surely has to **use** this structure to legitimate his/her product.

As far as the case of summarizing by the very producer of the original text, he/she occsionaly produces a summary without looking back the original text. This is clear in my data, which I will present at the seminar.

A3.
My answer depends on the definition of the "effectiveness" in the question. Effective for general readers and users of the summarizing system? Yes, to some extent. The system organizers may have to show very good rules to use them.

Effective for researchers and specialists who study the process of summarizing or who develop a better system? I think yes.

A4.
I think summarization is potentially in all our language activities. When we communicate with others, we do that by using language (in a broad sense). When we express what we try to let others know, we have to process an enormous amount of information to produce a text, which will serve our trial in interactional settings. This is what Dr. Aaron Cicourel has proposed in 1978. I think what he proposes is clear in my finding that the authors have to elaborate in selecting the most appropriate expressions to write not only summaries but also original texts.

As concerns the current summarising strategies that we have explored so far, however, we should rather study very specific domains or materials first, I think. Because we know just a few.

HS1.
Concerning the summrising by information specialists or abstractors, empirical studies have demonstrated that the process of summarising is task–specific so that certain trainings assure the considerable correspondence of the products among the trained persons.

RD1.
Does the question ask "human discourse processing strategies" that are illustrated in "discourse models" or "mental models"? I think Prof. Brigitte Endres–Niggemeyer have a good evidence for it. From my standpoint, discourses are recognized as "discourse (in general)" and "summary of the other discourse" in people's praxis of handling and classifying them.

RD2.
It depends on the settings where the consequent summaries are used. If they are written by the author when he/she submits a paper, they should be recognized as a good guide. When they are read quickly to grab the gist (of the original text), they should serve the readers to tell what they like to know. The organizers and developers of a certain information system need them to be precise in transferring information.

Woojin Paik

A1.
I would say no. Even the simplest existing text summarizing systems which work on narrow text type rely on the interpretation of text that summaries tend to occur at the beginning of texts. My opinion is based on the broad view of the text interpretation (not limited to the natural text understanding)

A2.
Yes, the focus or the main theme wich can be identified from the global text structure is the undoubtedly important information for the summarization task

A3.
It depends on how we define the notion of 'purely linguistic processing'. Since I include pragmatics as a sub–discipline of linguistics, I would say yes.

AS1.
Yes and no, there are certainly components of techniques which are independent from the text type, the subject domain, or the users. However, there are also components which are specific to the text type or the subject domain. However, if we look at each technique as a whole, I would say no.

AS2.
The answer is the same as AS1. It depends on whether you look at each technique as a whole or a collection of functions in the techniques.

Cecile Paris

A1.
Human processing most probably does not decouple interpretation from summarization. One must indeeed understand a text (at least to some level) in order to summarize it.

Computationally, however, summarization can (and probably should for ease of processing) be decoupled from interpretation. (I take interpretation here to mean parsing of a text). Summarization can be done from an internal representation (a text has already been parsed into an internal representation; alternatively, summarization can also be done on data obtained directly from a computer — no text interpretation. (number summarization; summary of a system's processing in expert system, for example). In these cases, summarization will include imposing a view, or interpretation, on this data in order to form a summary. This is not parsing however.

A2.
Summarising depends on the recognition of both the intention of the writer in writing the original text (with respect to what he or she was trying to convey) as well as the goals and knowledge of the reader (why do they want a summary and how much do they know abou the domain).

Insofar as the recognition of the global text structure helps determine the intention of the speaker, it is useful to summarisation. If it were possible to determine the intention of the speaker without the global text structure, then summarising would not need the global text structure either.

A3.
I am not sure what this question really means. First, we need to define "effective" as well as "purely linguistic processing" (i.e., what exactly is included in that processing). I believe that situational factors as well pragmatic knowledge are necessary for effective summarisation. (By effective, I mean that the summaries are accurate summaries of the text of data they summarise as well as they provide their reader with the information they needed and desired, in terms they can understand).

A4.
There are most probably general strategies as well as domain– or situation–specific ones, and it is likely that a combination of both is usually necessary. How general (and still useful!) can a strategy be must be studied through extensive experimentation.

HS.
Questions on human summarising: This is not my field and thus I cannot say much about the processes employed by professional abstracters. I believe they do go through training, though.

AS1.
Current techniques are application specific (both text type and subject domain — I do not know of any techniques oriented toward the user). Because of the need for pragmatic factors and world knowledge, I think current techniques are necessarily application specific at this point in time.

AS2.
Does this question mean: Can current techniques provide a summary from a large–scale text? I don't believe this is possible with current summarization techniques.

If, on the other hand, the question means: can a long summary (given a very long text, or a large set of internal data) be obtained given current techniques? It should be indeed possible. There are a number of linguistic theories regarding text structure that indeed can capture large–scale text. Producing a long summary whose structure is captured by one of these theories is thus possible.

Unfortunately, the state of the art in text interpretation is not quite yet such that these theories can be applied to the understanding of large–scale text in order to derive its structure.

RD1.
When attempting to do automatic summarisation, we often study human summarisation and try to duplicate it. As a result, human discourse strategies affect automatic summarising strategies.

RD2.
Topicality, nuclearity, discourse relations.

UA1.
As with any text generation, tailoring to the intended audience is important: the readers need to obtain the information they need in terms they can understand. As mentioned previously, I take an effective summary to mean a summary which not only accurately capture the information of the original text or of the internal data, but also take the user's goals and knoweldge into account.

Whether tailoring should be done at the individual or audience type level depends of the situation.

UA2.
Goals and interests in reading a summary; domain knowledge; linguistic competence (syntax as well as lexical)

There are a number of techniques developed in the field of generation and user modelling that are concerned with choosing the appropriate content, structure and phrasing of a text given a user model (whether individual or general). These techniques can be aplied to summarisation.

CR1.
They can help in that they can be used to provide information at the global text level.

CR2.
Template–based processing techniques can definately be used to find a template. This can be done with appropriate indexing. (may be I am not understanding the question, however).

Lisa Rau

A1.
Text summarization can and should be decoupled from text interpretation. Because text interpretation methods require large amounts of knowledge engineering for each and every domain to be understood, and still have yet to produce interpretations beyond extraction of simple factual information, we cannot expect that summaries based on conceptual understanding of text will be achievable in the near future. On the other hand, simple, domain–independent statistical and heuristic methods can perform very adequate summarization and should be the focus of research to produce automated methods today.

A2.
Summarizing absolutely depends on determining global text structure. Effective summaries can only be produced with systems that have some kind of understanding of where the important information in the text resides, and that can handle multiple different kinds of text types, with arbitrary internal structure.

A3.
I do not believe that effective summaries can yet be obtained with purely linguistic processing.

A4.
I believe that it is possible to produce very general–purpose summarising strategies.

HS1.
This varies very much with the nature and purpose of the text to be summarized.

HS2.
News are better than scientific articles which are better than novels.

AS1.
Current techniques are application–specific with respect to text type, but probably not within subject domain or user need.

AS2.
Current techniques do not capture large–scale text structure across a variety of text types with good accuracy.

RD1.
Don't know.

RD2.
If text is not generated but constructed from pre–existing texts, reference resolution is important to prevent the inclusion of dangling references in the summaries. It is also key to identify the number and location of separate topics in texts, and what is introduction, subject matter and conclusion.

UA1.
User adaptation is a minor issue when the base functionality of summarization either from an indicative or informative perspective has yet to be achieved.

UA2.
Don't know.

CR1.
I have not found any domain–independent sentence processing method that helps to identify significant text content. Statistical methods do however help.

CR2.
Present template–based processing methods work very well in constrained domains to find as well as fill templates.

Ellen Riloff

UA1.
In many applications, I think that it is very important for summaries to focus on the interests of the user. For example, in information retrieval and hypertext scenarios, users often search through large document collections to find texts that are relevant to their needs. There are several reasons why tailoring summaries is especially important in these scenarios.

First, the user may browse through large numbers of texts. In this case, the user does not have time to read long summaries. Also, the user is probably looking for texts about a specific topic and does not care about information that is not relevant to that topic. In order to efficiently look at many texts, it is important for a system to generate short summaries that include only the information that is most relevant to the user's interests.

Second, large document collections often contain many texts that are long and discuss a variety of subjects. For these texts, a short summary of the entire document may be too general to be useful. For example, a general summary of a stock market report might read: "The stock market gained 100 points today in heavy trading. In general, car manufacturers did well but pharmaceutical companies lost ground." However, a user might only be interested in tracking the performance of a specific car manufacturer, such as Chrysler. A concise summary of the entire document would not necessarily be able to specify individual items that are of interest to the user.

UA2.
In general, it is often difficult for users to explicitly list features that are useful for recognizing texts that are relevant to their interests. A user can easily forget or overlook good features, or inadvertantly include features that are potentially misleading or not discriminating enough. One approach to this problem is to use statistical techniques and large corpora to automatically identify good features.

The advantage of statistical techniques is that they can automatically sift through many examples of what a user has deemed to be relevant and irrelevant in the past. Given a training corpus of texts that have been manually classified as relevant or irrelevant by a particular user, a system can exploit the corpus to automatically identify features that represent the user's interests. Given a new text, the system can then use these features to select information that should be included in a summary tailored for the user.

Lucia Rino

A1.
Yes, text summarising can be decoupled from text interpretation.

A2.
If we consider a summary as an expression of the content of a text (i.e. informative), the answer is yes. The summary would be based on a global text structure and would correspond to a sub–structure of the text. A global text structure should also underpin the construction of indicative summaries. In both cases, the global structure will allow the proper organisation of relevant information and the proper condensation of content according to the audience.

A3.
No. Linguistic processing comes after the proper structuring of summaries, for the selection of surface information. Without a coherent organisation of information, it is not possible to assure that linguistic decisions will provide the means for effective summaries.

A4.
Considering that summaries can be classified according to their functionality, which comprises coverage, degree of informativeness, selectivity of information, and finally ways the readers are addressed, there is a domain–specific and a goal–oriented order which will be expressed in discourse strategies. For this reason, it is quite difficult to think that summarising strategies can be general.

HS1.
I have no idea, but humans rely on training to write well.

HS2.
It seems important to consider the way humans scrutinise the text in order to select the most relevant information and related propositions, their control over the clues given by markers, or selection of topic information, and the way they decide about theme development vs. relevance. In short: the heuristics they use to select, organise and express the information extracted from the text.

AS1.
Yes.

AS2.
It is not clear to me what is considered here to be a large–scale text structure.

RD1.
I don't understand this question.

RD2.
Discourse strategies to organise coherently the information and properly address the audience.

UA1.
Very important to tailor, but not always necessary to tailor for individual readers.

UA2.

User features:

world and linguistic knowledge, goals (which will determine the relevant information to be conveyed into text, i.e. relevance assessment).

Tailoring techniques:

goals, choice of relevant information (i.e. topic information which is crucial for a specific readership), amount of detail (explicitness for the reader), linguistic means to express the underlying structure of the text.
They are related to:

- generalisation/specification techniques,
- selection of information,
- specification of the language for the special purpose of assessing the reader appropriately,
- rhetorical techniques to address the reader.

CR1.
Since sentence processing methods lack the overall conceptualisation given by the text, it cannot, on its own, provide all the significant information. For instance, the identification of significant text content is related to thematic progression, so that coherent spans of text must allow the inter–sentential link to be captured from the text. If we consider just sentence processing methods, the overall organisation and relationship between units of information is missing.

CR2.
I would imagine so.

Annely Rothkegel

A1.
It is not clear what "interpretation" means. In case it refers to categories of knowledge about the particular domain of the original document it is assumed that some understanding is necessary in order to select the relevant pieces of information.

A2.
According to the thesis of my talk it does! The reason is that summarizing can not be understood in the sense of reducing a text (the document) but of producing a text (the abstract). In order to produce a text it is necessary to have a linguistic knowledge about global text structures (thematic development, completion of the goal of the text).

A3.
cf. A.1; I think that also knowledge about the domain is necessary. (It may be possible to produce an abstract without any knowledge of this kind. The question whether the result will be a "good" abstract.)

A4.
With respect to the linguistic knowledge the strategies may be very general.

HS2.
If we distinguish between two levels of abstracting – the forming of the text structure and text on the one hand, and the phases in which the writer is involved on the other hand, then I think that the text related strategies are the best candidates for automation.

AS1.
I think that there are some general techniques as they are used for purposes of text production.

RD1.
They are the basic strategies for summarizing.

RD2.
The construction of coherence.

CR1.
It is not a matter of sentence processing!

Donia Scott

A1.
It is difficult to see how human summarising could be decoupled from interpretation. Interpretation comes into play in two ways:

1. The summary–writer's interpretation of the text.
2. The interpretation the summary–reader is expected to have.

For automatic summarisation, they can (and for greater ease, probably should) be viewed as two different tasks. For example, summaries can be generated from an internal representation that does not depend crucially on interpretation.

A2.
Summarisation does not necessarily depend on the recognition of global text structure, but rather on the intention of the writer. Access to the global text structure can, however, facilitate the process of summarisation.

A3.
Not having a clear definition of "effective" and "purely linguistic", let us simply say that world knowledge is definitely required.

A4.
This is an empirical question which requires substantial experimentation. One would imagine that summarisation is best achieved through the use of at least some general purpose techniques, refined with domain/audience specific ones.

HS1.
HS2.
I am not familiar with the processes employed by professional abstractors.

AS1.
Current techniques are both domain– and text–type– specific. I am not aware of techniques applied to user–need. Given that most current techniques do involve interpretation, the tradeoff between power and generality not only cannot be avoided, but is heavily stacked in favour of application specificity.

AS2.
As far as I am aware, current techniques do not capture large–scale text structure.

RD1.
Automatic summarisation is clearly related to human discourse processing in that a weak–replication of human processing strategies is attempted.

RD2.
Topic, discourse relations, nuclearity.

UA1.
Tailoring is clearly important; whether it is to a general audience or an individual user depends on the situation. Tailoring addresses the needs of the intended reader(s): their goals and interests in reading a summary and their level of domain knowledge. Presumably an "effective" summary is one that takes these into account.

UA2.
The features outlined in the answer to UA1, in addition to (as with any text) linguistic competence. Tailoring techniques include the user–oriented content–, structure–, and phrasing– selection.

CR1.
Only to the extent that information obtained at the sentence level can contribute to information at the text level.

CR2.
Surely template–based processing can be used to retrieve templates, given appropriate property indexing.

Karen Sparck Jones

Computer Laboratory, University of Cambridge

A1.
Of course summarising requires *some* source text interpretation, eg at the sentence level. For summarising as an explicit task, in sufficiently well–specified circumstances with eg predefined information extraction needs, interpretation may be direct to the summary representation (even text). Otherwise a full source representation is required. In reading text or listening to spoken discourse, summarising is an intrinsic part of the interpretation process, building a rolling summary representation.

A2.
For text summarising in general, it is necessary to capture global text structure but, beyond the rock–bottom structure defined by a statistically–motivated keyword list, it is not clear what this structure is or how it should be used. Much more experiment with different ways of characterising and using global text structure is needed, for instance is it derived or recognised? It is also not clear how far structure has to be explicitly captured and manipulated to obtain a summary as opposed to merely exploited (as in taking news story opening sentences because they happen to contain important content).

A3.
It depends what is meant by purely linguistic processing, since a good deal of domain knowledge can be smuggled in via the lexicon and semantic selection restrictions and thus be reflected in the output, say as logical forms, of sentence analysis. Whether useful general–purpose summarising, as opposed to extracting, can be done by such purely linguistic means has to be properly tested, especially for long sources.

A4.
The limits on general–purpose summarising come from intended audience constraints. A general–purpose summarising program has therefore either simply to respond to a source text as written for its original audience so, eg, if the source was for some specialised audience the summary will be for the same kind of audience; or assume a 'standard' audience to which summarising is uniformly addressed; or it is necessary to generate different summaries explicitly for different audiences.

HS1.
Any language user can and does summarise adequately (even effectively) for their immediate needs: training is needed for longer–term value.

HS2.
Interesting question how common professional practice of cycling (draft, revise, finalise) can be automated.

AS1.
Given the tradeoff between D = deep but narrow and S = wide but shallow, current D methods are application specific, S ones need not be.

AS2.
Current methods pretty well don't capture large–scale structure, especially for long texts, except for very limited statistical or extractive approaches.

RD1.
Human continuous content consolidation is relevant, if we knew how to do it.

RD2.
Large scale structure is *very* important: the issue is what (kind of) structure and how to identify and use it.

UA1.
Tailoring to the user is fine with enough information to do it properly; but typically there is not enough information about the individual user, so a conservative approach is essential (cf, similarly, question answering, document retrieval, document routing). Information search and use is a dynamic, interactive process, so a summary makes only a partial contribution in any situation.

UA2.
User feedback can be helpful (but getting, applying, and generalising from it is a problem with one–off contexts).

CR1.
Sentence processing can help via eg content decomposition into logical forms; but we need more work on what rich lexical information can deliver for summarising so eg we can use sentence representations effectively for modelling content relationships lurking in anaphoric references and in local cohesion relationships (eg, say, elaboration).

CR2.
Template methods are a good field for experiment, ie how to develop, modify initial a priori templates under extensive text sampling.

Hans Strohner and Gerd Rickheit

A1.
No, because text summarizing includes not only one, but two types of text interpretation:

- text interpretation of the original text in order to transform it into a summary,
- text interpretation of the summary.

A2.
No, because event structure is more important than text structure.

A3.
No, see question A2.

A4.
Summarizing strategies have to include at least sensory–motor, syntactic, code, reference, semantic sense, and pragmatic strategies, as outlined in the paper by Strohner & Rickheit.

HS1.
Summary authors should master the strategies mentioned in A4.

HS2.
Sensory–motor, syntactic, and conceptual aspects.

RD1.
Summarizing is a type of discourse processing.

RD2.
Reference, semantic sense, and pragmatics.

UA1.
Very important, because the reference, semantic sense, and pragmatic inferences in processing summaries are highly different between individuals.

UA2.
Experimental research in summary processing.

Patrizia Violi

A1.
I don't think it is possible to decouple human summarizing from interpretation. To make a summary we have first to understand a text, and understanding always implies interpretation. However I believe a distinction can be made between the two activities. Interpretation is a less constrained process than summarising. Summary, as a textual genre, has some specific constraints, for example it should respect the relevance hierarchy present in the source text (if any), while interpretation can, and indeed should, inlight more "hidden" aspects of a text. Computationally I am not sure what it means for a computer to "interpret". If it means to parse, then the two issues can be decoupled. If it means "understand" I do not believe computers understand, only people do, then we are back to human summarising. (But this is probably a different workshop).

A2.
In general terms, and always thinking of human summarising, I think the answer is yes: in order to summarise we need to grasp a global text structure.

A3.
It is not clear to me what is meant by "linguistic" processing Obviously we need world knowledge to understand texts. (Indeed, I doubt it's even possible to make any principled distinction between linguistic knowledge and world knowledge).

A4.
I have no ideas.

RD2.
Certainly an understanding of the topic, of the global structure and the relations among components.

Outcome of the Seminar

A Research Platform for Intelligent Summarizing I

Friday 17–12–1993
9.00 – 10.30

Chairperson: Jerry Hobbs
Rapporteur: Brigitte Endres–Niggemeyer

The discussion was organized along the following list of seven questions. However, not all questions really stimulated reactions. The questions that were not commented upon are skipped in the report.

1. What is a summary? What is its source and its relation to its source?
2. Ideally, what is required for constructing a good summary? Text structure, genre, world knowledge, author's intentions, other text parameters?
3. What are the uses/types of summaries? Is there a "use–neutral" notion of summary? What can we do with current shallow techniques? Physical structure, statistical identification of key sentences/concepts, information extraction techniques, compacting information? What near–term applications are there for current techniques? Are abstracts passe, with interactive systems?
4. What can we do with current shallow techniques? Physical structure, statistical identification of key sentences/concepts, information extraction techniques, compacting information?
5. What near–term applications are there for current techniques?
6. Are abstracts passe, with interactive systems?
7. How to evaluate summaries,e.g., acceptability, readability, coherence, cohesion? Are there cultural differences?

1. What is a summary? What is its source and its relation to its source?

and

3. What are the uses/types of summaries? Is there a "use–neutral" notion of summary?

Summaries seem hard to define: Participants mentioned that

- they are shorter than the source text
- they contain less information
- they may convey explanations missing in the original and thus exceed the original document

in length
- they reduce information to the essential items
- they save reading time
- they are learning tools

One can define the concept "summary" by a set of features according to its uses, without insisting on a common semantic core. The question is if a "use–neutral" partial definition of an abstract is possible. Since the use of a summary or abstract influences its definition, the "use– neutral" definition would be augmented by features won from its uses.

The relation between a summary and its source gave rise to controversial opinions. One may characterize the relation as indicative, informative, or evaluative. While some speakers were in favour of keeping these classical distinctions, it was asked what is meant by "informative", the difference between informative and indicative was judged old–fashioned, and one speaker proposed to describe summaries and abstracts by a set of facets describing the type of information, and to explain what is "informative" or "indicative" by means of these facets.

2. Ideally, what is required for constructing a good summary? Text structure, genre, world knowledge, author's intentions, other text parameters?

To construct a good summary one needs

- techniques for determining what is in the source
- techniques for building the summary

Some subtasks of analysis are easier than others, e.g., it is less hard to find out about the text structure than about the author's intentions.

In summarizing, there is no natural sequence of source interpretation and generation. In particular, it seems often more practical to analyse the source text from a generation perspective, i.e. to find a concrete piece of information that is needed for summary generation.

Both sentence syntax and discourse structure contribute to automatic summarizing of language text. In addition, it is necessary to account for non–linguistic, e.g. multimodal summarizing.

User–oriented tailoring of summaries and abstracts can recycle ideas known from question answering: A summary is then constructed in order to answer a user's question.

Fig.1 gives the overall image of summarizing. Resources involved are listed in its first version, whereas the second one shows where current systems and approaches have put their main emphasis.

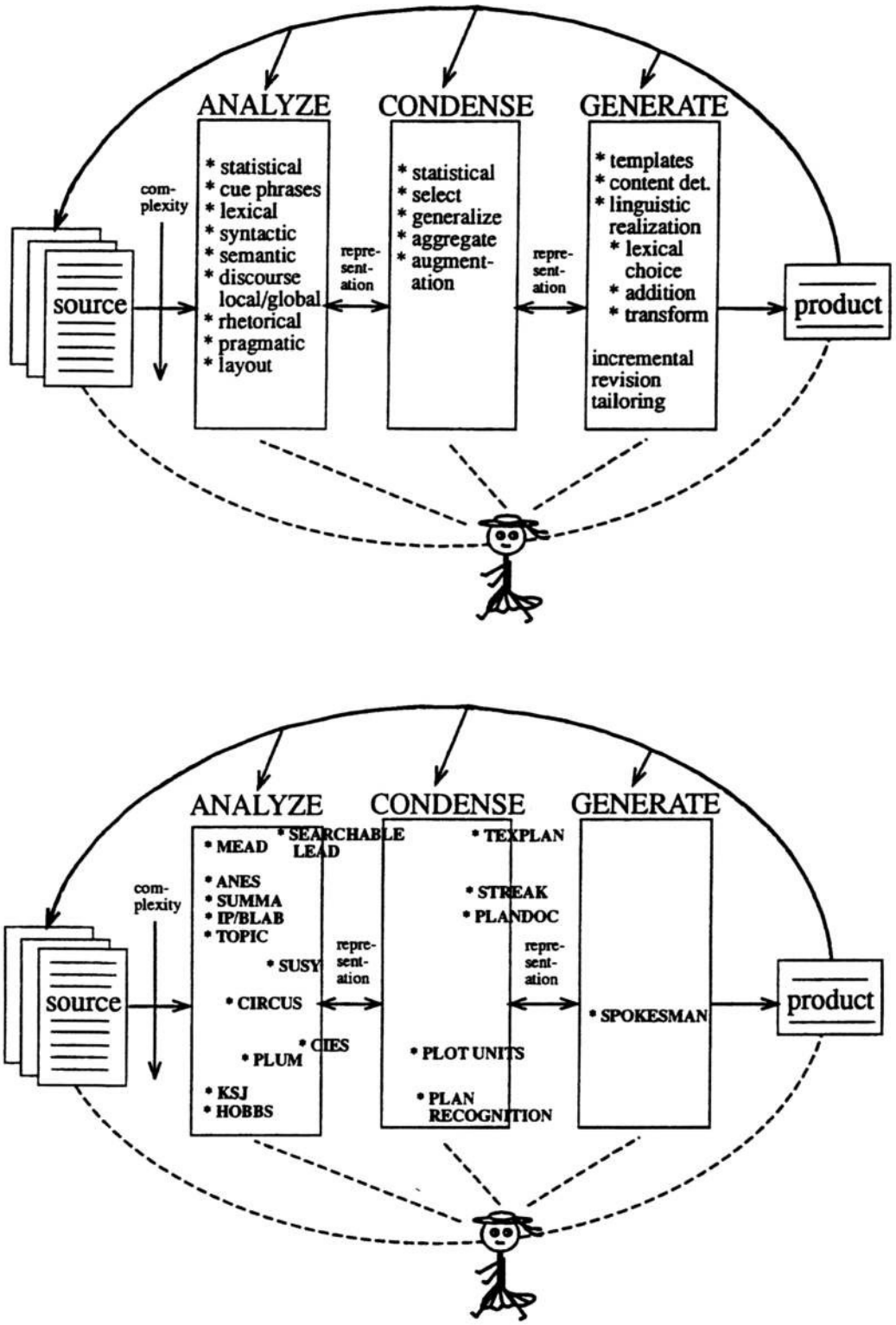

Fig.1: Research in summarizing:
Above: Resources involved,
below: Main emphasis of some current systems and approaches

7. How to evaluate summaries, e.g., acceptability, readability, coherence, cohesion? Are there cultural differences?

It is just as hard to evaluate summaries to decide what is a good text in general. Since one may want to judge how much a system's summaries have improved, the problem of evaluating success adds to the problems of text assessment. The purpose of the summary is important, a classification of inputs and purposes is helpful.

Evaluation techniques can be intrinsic, looking at summaries, e.g., comparing them with guidelines or source texts; and extrinsic, i.e. assessing the abstract's fitness for a particular task.

Both experimental and naturalistic (field work) approaches seem possible. It is important to achieve more than a mere assessment in terms of traditional recall and precision measurements.

A Research Platform for Intelligent Summarizing II

A research agenda for automated summarizing

Friday 17–12–1993
11.00 – 12.30

Chairperson: Karen Sparck Jones
Rapporteur: Brigitte Endres–Niggemeyer

Overview:

1. The starting point of the seminar
2. Preliminary use of the question list
3. A sketch of the current research situation
4. Setting up a research agenda
5. Practical hints

1. The starting point of the seminar

A descriptive framework for summarizing

The discussion of the seminar was guided by a descriptive framework for summarizing proposed by Karen Sparck Jones. It comprises three main classes of factors: input, purpose, and style (see fig.1). The question is how to combine input and purpose to constrain output.

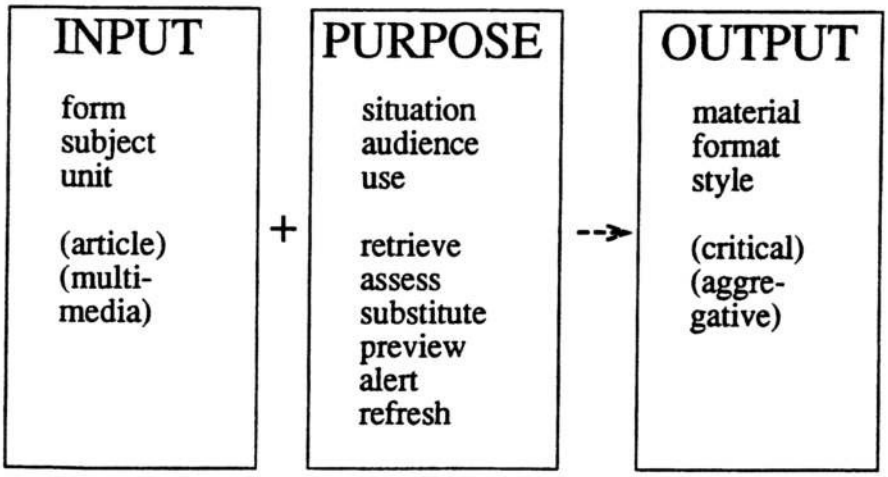

Fig.1: Factors of summarizing

We get the process structure of fig.2. There, all representations except the source text are the (intermediate) products of the component processes of summarizing.

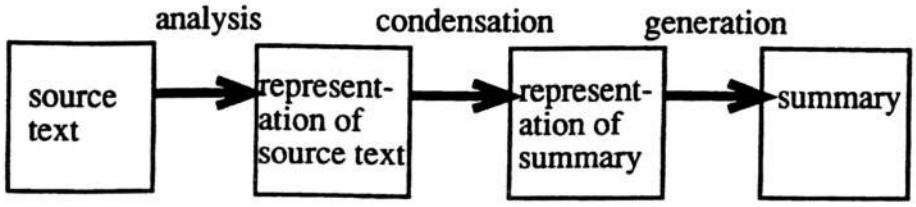

Fig. 2: Process structure of summarizing

Issues and goals of research

We started the seminar with a list of issues that might guide our investigation of summarizing:

- the properties of the source representation (text, database) that enable to summarize
- the way action–oriented accounts of summarizing can be integrated with text–oriented ones
- the influences of contexts: types of input materials, specific purposes (genres of text, properties of users, and of what they use summaries for, and so on)
- the importance of deep vs. shallow strategies
- the evaluation of summarizing

Wolfgang Wahlster proposed strategic goals in order to achieve a more comprehensive scientific treatment of summarization:

- a formal definition of an optimal summary
- an assessment of the relative qualities of summaries vs. extracts
- roles for textual, multimodal, or interactive summaries
- the representations required for different summaries
- the nature of the resource limitation, and resource limited techniques
- evaluation methods

Research for automated summarizing may have to be conducted with respect to human summarizing performance.

2. Preliminary use of the question list

For several reasons, only a preliminary discussion of the question list was possible. An educated guess of the participants' opinions as expressed orally during the seminar turned out to be most interesting when uncovering hidden ambiguities and controversial views. The questions A1 and HS1 stimulated most discussion:

A1. Can and should text summarizing be decoupled from text interpretation?

This question may read as "Should one build a source representation before thinking about summarization?" Then the answer is that source representations can be used to decouple text interpretation and summarizing. The model of the text can be independent of the summary representation.

However, the text interpretation that sets up the source representation is part of the later summarization achievement. This means that even through a sequence of representations, the final summary will remain influenced by the interpretation of the source text. No real decoupling takes place.

HS1. How far does human summarizing depend on task-specific training?

Summarizing is a part of normal human language processing. Therefore, a basic competence in summarizing would not need task-specific training.

There are, however, important variations in human summarizing skills and summary quality. They are due to task-specific training. Consequently, human summarization relies heavily on task-specific training as soon as more than basic skills are considered.

3. A sketch of the current research situation

Only a part of the concepts that seem useful for automatic summarizing have been investigated rather thoroughly. Fig.3 places the better investigated concepts on top. Concepts that found scarce or no attention in research figure at the bottom.

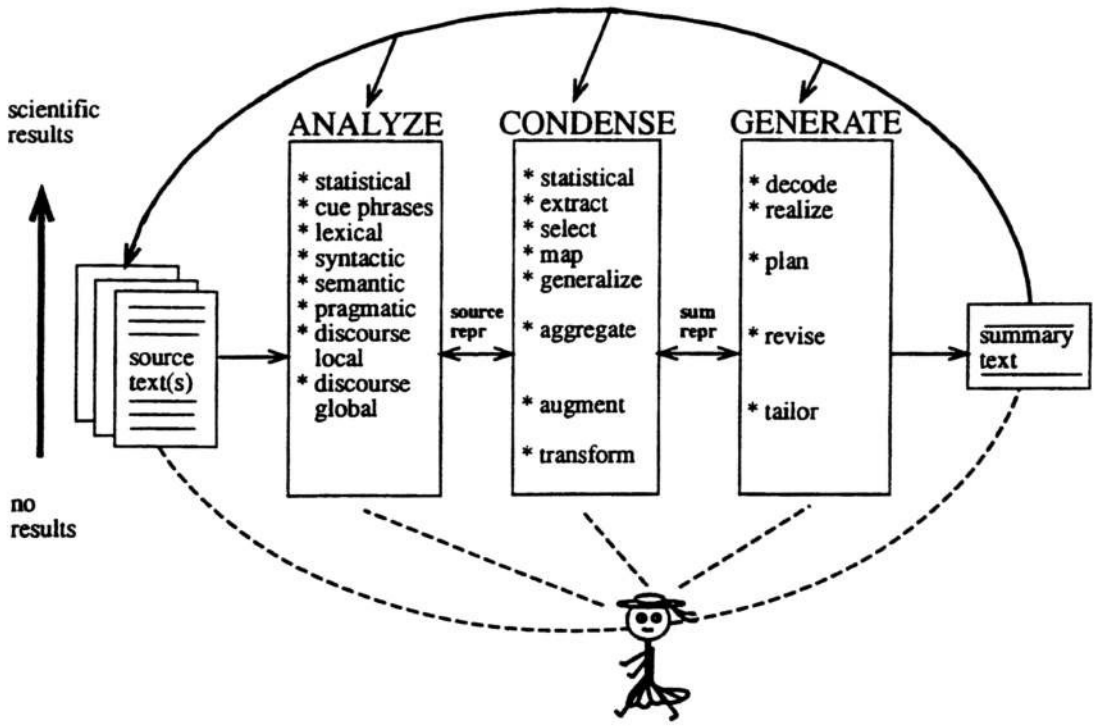

Fig.3: What research has been done in automated summarizing (bottom: scarce activity, top: substantial research, feedback loops shown)

4. Setting up a research agenda

According to what we have learned during the discussions of the week, a research agenda for summarizing should focus on uses and users of summaries (see fig.4).

Uses of summaries are manifold: retrieval, assessment of documents or the database, preview, substitute of the original document, refreshing, alerting, and so on.

By focusing on uses and users, we can decompose the broad factors in input, purpose, and output (see above) into features. For instance, if one wants a summary to refresh knowledge, or to preview a later complete reading of the original text, one can decide what kind of properties the summary needs in order to support the chosen function.

It is important to concentrate on the points that make summarization unique: condensation with subfunctions like generalisation and aggregation. This is the more true since they happen to be among the least explored areas of text processing.

From the lack of existing research in condensation follows the need to develop both principles and technologies. We have to ask what it means to develop new automatic methods, instead of providing ad–hoc methods without a valid functional interpretation. Meaningful methods can be attained more easily by approaching summarization through its uses.

In addition, we have to identify the influences on analysis and generation which are specific to summarization. The question is how current analysis or generation components must be changed when integrated into a summarization system.

Since summarization happens in a rapidly changing media environment, we have to deal with new forms of texts and summaries, e.g., with multimodal or interactive documents and summaries.

The research agenda should comprise the following actions. They may be investigated in parallel (after the first ones):

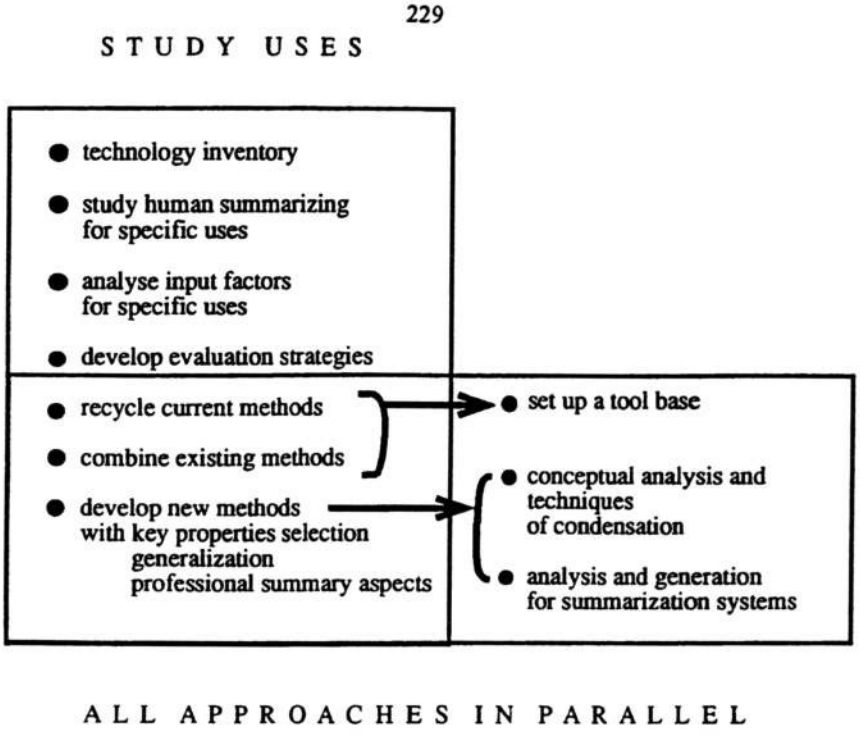

Fig.4: The research agenda

The technology inventory

A technology inventory or tool base for automatic summarizing should start with existing automated methods. This should be enhanced with concepts that are not yet implemented, but belong to the conceptual apparatus of competent human summarizers.

Studying human summarizing

When studying human summarizing, we have to focus on summarizing for different uses and users, on what human summarizers or abstractors are doing when summarizing for a particular use.

Analyzing input factors for particular uses

Input factors as subject, domain, jargon, what kind of audience the original document was addressing, or genres all influence the derivation of an abstract for a certain task. Up till now, research has rather wrapped up these factors as features of text, instead of tracing their individual consequences systematically. We have to ask which effects these features have on summarizing for a particular use.

Evaluation strategies for automatic summaries

Evaluation can follow two models:

- Intrinsic evaluation, i.e., asking, e.g., how well the summary represents the source text, about its cohesion, its adherence to guidelines, or its similarity to human summaries which are assumedly always better than automatic ones. This evalation type has its merits, without appearing completely satisfactory.
- Extrinsic evaluation based on summary use. A summary is as good as the level of performance a user using it can reach with its help. For this type of evaluation, experimental and naturalistic methodologies can be developed or adapted from information retrieval. But since summaries are much richer objects than queries, their use risks to be much richer, too, and their evaluation more demanding.

Recycling current methods

We need to find out more about known technologies, about what we can do with them. Even the rather crude extraction strategies may be adequate in some contexts (cf. the success of machine translation systems like Systran or Metal). If so, existing methods can be recycled for specific uses. Tool sets or databases are useful for both recycling and combination (see below) of known methods.

The combination of known methods

For specific uses, existing methods may be more appropriate when combined in a productive way. In order to allow for the combination of existing methods and systems, we have to develop interfaces and representation languages for defining, e.g., document structures.

Developing new methods

The development of new summarizing methods should concentrate on the condensation subtask because captures the kernel functionality of summarizing best, and because we lack know–how precisely about this core task of summarizing. Both concepts and techniques must be developed.

5. Practical hints

We set up a common bibliography of the scattered summarization research from the different disciplines. Please send your references to Brigitte Endres–Niggemeyer.

Ed Cremmins offers to procure summarizing literature from the Library of Congress and other libraries in and around Washington DC.

The organizers' conclusions

Dagstuhl works like a scientific greenhouse. There, a week's discussion on automatic summarization led us to a research agenda which invites – among other things – to combine existing summarization tools, to develop evaluation strategies, to study human performance, to conceive new implementable summarization concepts and techniques, and to consider summarization not as a uniform activity, but as a flexible process adapting to context factors as the intended use of the summary.

The soft greenhouse climate eases communication and new insights. If these insights are viable is decided outside, in the wide landscapes of cognitive science, information systems research, AI, and computational linguistics. Therefore the following sketch highlights macroscopic features of the scientific and technical background that we had in mind during our discussions and explains how the discussions of this report integrate with the overall research field. On the whole, we observe an increasing demand for automatic summarizing, for more diversified and sophisticated summarization products and styles, the growing affinity of automatic and human summarizing research, and a need for better scientific methods.

1. Larger systems increase the demand for summarization

In information systems of the future, full text data bases, distributed data bases, integrated multimedia and hypertext information systems will all contribute to present users with more material. Since users cannot expand their information absorption capability correspondingly, summarization will be more necessary than ever to keep information sizes manageable. In addition to their basic alerting and reference function, summaries must help systematically to make an intelligent use of information, for instance by previewing larger items (documents, or document collections) and by organizing informational contexts for users venturing into more or full informational detail. With increasing information sizes, summarization becomes not only more important, it gains additional functions as well.

2. Multimedia communication and emphasis on summarizing principles

In the same time, the arrival of interactive hypermedia and multimedia frees automatic summaries from the ASCII characters constraint that imposed the format of written language text. In current systems, both computerized documents and their summaries may be multimedia texts, including images or even a sound track.

Since media (image, text, graphics, sound, etc.) can be chosen in order to render the informational content most clearly and concretely, summaries, e.g., scientific abstracts, risk to change their

appearence dramatically. Times are gone when a summary may have appeared as a "Ding an sich". Today's researchers in automatic summarization can start out from text processing and choose additional media deliberately where their forefathers were bound to written text.

The technological shift to multimedia environments affects not only the outer appearance of summaries, but also the structure of automatic summarizing systems. As the presentational form of a summary becomes a matter of choice, the summarization process wins more independence from its source and its product. Please imagine a summarization system using ad–hoc surface cues of different media only, instead of summarizing content in a media independent object representation. Without such a representation, the system would neither be able to integrate information from different media, for instance, from an image and a textual statement, nor would it be able to allocate information intelligently to appropriate media. Presentational flexibility makes it harder to achieve defendable summarization–like results by shallow or ersatz operations, without applying real meaningful summarization concepts. By doing so, it supports a more principled approach to automatic summarization, and a decoupling of the working process from the surface features of both input and output media.

Multimedial presentation is currently the most conspicuous advance in computer communication to influence summarizing, but by no means the only one: In particular, interactive systems invite to conceive summarization as an interactive process instead of a simple display of summaries. The consequences for the design of interactive summarization systems are just as evident as for systems that would tailor summaries to particular uses: These systems will need more theoretical underpinnings than earlier solutions.

3. Coming nearer to human summarizing functionality

As technical development frees summaries from presentational restrictions, summarization systems need more intelligence to cope with the additional degrees of freedom. Like this, technical advance presses for a more articulated and penetrating view of automatic summarization, and in the same time it reduces the conceptual distance between automatic and human summarizing. Where earlier researchers in automatic summarization saw no chance to learn something implementable from the overly complex skills of human summarizers, their colleagues of today acknowledge that automatic summarizing has to be defined with respect to human performance. After all, competent human summarizers dispose of the most elaborated summarizing competence available, and more intelligent summarization concepts can be won naturally by studying human skills. Consequently, cognitive approaches have more impact on system design than in former times.

A basic observation tells us that cognitive systems like human summarizers do not summarize texts, but states and events in the real world which may be reconstructed from a written text or an image sequence by understanding. When the task is to summarize from a text, discourse features

are helpful, but the condensation process essentially reworks the object representation built up by text understanding and derives a summary representation that is uttered as spoken or written discourse. System design can follow the human process architecture.

4. Towards a science of summarizing: More empirical foundations, more formal accounts, more interdisciplinarity

Although systems that realize a shallow concept of summarization may be satisfactory in limited environments, real summarizing is a complicated intellectual process where many pieces of knowledge are combined in an organized way. Unfortunately, we lack concrete knowledge about many of its features. The current concept of a summary, for instance, describes a conciser statement of the most important items of a larger unit like a text. For use in automatic summarization, it appears poor. Indeed, more meaningful summarization theories are needed for more sophisticated systems.

Even though formal definitions of a summary or of summarization seem out of reach, it is desirable and feasible to provide implementable accounts of summarization which are more organized, more complete and empirically and formally more valid than approaches of the past. This is in part a matter of methodology. Summarization research is still improving its scientific methods, e.g., by imports from related disciplines like information retrieval. Most interesting are methods that allow a modularization of summarizing, implementing a holistic "divide and conquer" approach aimed at empirically and formally well–founded accounts of interesting features like subtasks, or special types of summaries. New concepts may be won by formal (deductive), by constructive and by empirical procedures.

Since summarization is defined by the related human skills and concepts, cognitive approaches help to establish how a summarization process is organized, which features of the source text influence the resulting summary, how intended uses shape the summary, and so on. Evaluation methods are just as important as empirically founded constructive methods.

Because knowledge about different types of objects (discourse types, cognitive processes, representation formalisms, etc.) must be combined to explain summarization, successful research is hard to imagine without interdisciplinary cooperation.

Most interestingly, modular theories of summarization allow to single out the specific condensation processes of summarizing, and to separate them in particular from general purpose text understanding. Like this, summarization research can concentrate on the functional kernel of summarizing. In a second move, general text understanding and production components can be adapted to summarization systems. Basically the same modularization strategy can serve to enrich our knowledge of texts and summaries. By subspecifying source documents and summaries with specific features, e.g., drawn from uses like "an alerting summary for social scientists", one can improve definitions that are too pale to guide production and to support evaluation.

Address List of Participants

Status of 19.2.94 with updates for Bateman, Endres–Niggemeyer, Rothkegel, and Sigel

[**Bateman, John A.**
GMD-IPSI
Dolivostr. 15
D-64372 Darmstadt
Germany

Nicholas Belkin
Rutgers University
School of Communication
Information and Library Studies
Huntington Street
New Brunswick NJ 08903
USA
belkin@cs.rutgers.edu
tel.: +1 (908) 932-8585

William Black
University of Manchester
Institute of Science and Technology
CCL
Sackville Street
Manchester M60 1QD
Great Britain
bill@ccl.umist.ac.uk
tel.: +44 (61) 200-30 96

Harold Borko
Univ. of California at Los Angeles
Graduate School of Library and
Information Science
405 Hilgard Avenue
Los Angeles CA 900241520
USA
tel.: +1 (213) 825 1379

Bruce Britton
University of Georgia
Institute for Behavioral Research
Athens GA 30602

USA
ibr@uga.bitnet
tel.: +01 (706) 542-18 06

Ines Busch-Lauer
Universität Leipzig
Fachsprachenzentrum
Augustusplatz 10
D-04109 Leipzig
Germany
tel.: +49-341-719-2961 (montags 15:30 - 16:00) /2977
fax: +49-341-209-325

Fabio Ciravegna
Istituto per la Ricerca Scientifica
e Technologica
I-38050 Povo Trento
Italy
cirave@irst.it
tel.: +39-461-31 43 15

Edward Cremmins
10075-6 Windstream Drive
Columbia MD 21044
USA
tel.: +1-410-740-8361
karla@jhunix.hfc.jhu.edu

Brigitte Endres-Niggemeyer
FH Hannover
Fachbereich IK
Hanomagstr. 8
D-30453 Hannover
Germany
ben@www.bid.fh-hannover.de
tel.: +49 (511) 44 43 44

Raya Fidel
University of Washington
GSLIS FM-30
Seattle WA 98195
USA
fidelr@u.wahington.edu
tel.: +1 (206) 543-1888

Helmut Felix Friedrich
Deutsches Institut für Fernstudien
Postfach 1569
D-72005 Tübingen
Germany
tel.: +49-7071-979-221

Rosemarie Gläser
Universität Leipzig
Fachsprachenzentrum
Augustusplatz 10
D-04109 Leipzig
Germany
tel.: +49-341- 19-2964 /2977 /2982 (Do. v. 10.45 - 12.00)

Udo Hahn
Universität Freiburg
Arbeitsgruppe Linguistische Informatik/
Computerlinguistik (CLIF)
Werthmannplatz
D-79085 Freiburg
Germany
hahn@coling.uni-freiburg.de
tel.: +49-761-203-32 55

Jerry R. Hobbs
SRI International
333 Ravenswood Ave.
Menlo Park CA 94025
USA
hobbs@ai.sri.com
tel.: +1-415 8 59 22 29

Paul Jones
Napier University
Department of Computer Studies
219 Colinton Rd
Edinburgh ED14 1DJ
Great Britain
paulj@dcs.napier.ac.uk
tel.: +44-31-455-46 77

Elizabeth D. Liddy
Syracuse University
School of Information Studies
4-206 Center for Science and Technology
Syracuse NY 13244-4100
USA

liddy@suvm.bitnet
tel.: +1 (315) 443-2911

Elisabeth Maier
DKFI Saarbrücken
Stuhlsatzenhausweg 3
D-66 123 Saarbrücken
maier@dfki.uni-sb.de
tel.: +49 (681) 302-5252

Mark T. Maybury
The MITRE Corporation
Artificial Intelligence Center
202 Burlington Road
Bedford MA 01730
USA
maybury@linus.mitre.org
tel.: +1 617 271 7230

Kathleen R. McKeown
Columbia University
Department of Computer Science
450 Computer Science Building
500 West 120th Street
New York NY 10027
USA
Mckeown@cs.columbia.edu
tel.: +1-212-939-7118

Sumiko Mushakoji
Tsukuba Science City
University of Library and Information Science
1-2, Kasuga, Tsukuba-shi
Ibaraki-ken
305 Japan
sumiko@ulis.ac.jp
tel.: +81 (298) 52 0511
fax.:+81 (298) 52 4326

Woojin Paik
Syracuse University
School of Information Studies
4-297 Center for Science & Technology
Syracuse NY 13244-410
USA
wjpaik@mailbox.syr.edu
tel.: +1-315-443-29 11

Cécile Paris
IT Research Institute
University of Brighton
Lewes Road
Brighton BN2 4AT
UK
cecile.paris@itri.bton.ac.uk
tel.: ++44+273-642900
fax.: ++44+273-606653

Livia Polanyi
Rice University
Dept. of Linguistics
P.O. Box 1892
Houston TX
USA
polanyi@ricevm1.rice.edu
tel.: +1-415-969-26 36 (home)

Lisa F. Rau
University of Pennsylvania
Dept. of Computer and Information Science
220 South 33rd Street
Philadelphia PA 19104- 6389
USA
lrau@cis.upenn.edu
tel.: +1-215-573-28 15

Ellen Riloff
University of Massachusetts
Department of Computer Science
Graduate Research Center
Amherst MA 01003
USA
riloff@cs.umass.edu
tel.: +1-413-545-24 40

Lucia Rino
University of Brighton
IT Research Institute
Lewes Road
Brighton BN2 4AT
Great Britain
lucia.rino@itri.bton.ac.uk

Annely Rothkegel
Hannover Polytechnic
Technical Writing Program
Bernhard Caspar Str. 7
D-30453 Hannover
Germany
rothkegel@coli.uni-sb.de
tel.: +49-511-212-3930

Donia R. Scott
University of Brighton
IT Research Institute
Lewes Road
Brighton BN2 4AT
Great Britain
donia.scott@itri.bton.ac.uk
tel.: +44+273-642900

Karen Sparck Jones
Cambridge University
Computer Laboratory
Pembroke Street
Cambridge CB2 3QG
Great Britain
ksj@computer-lab.cambridge.ac.uk
tel.: +44-223-33-4631 /4607

Hans Strohner
Universität Bielefeld
Fakultät für Literaturwissenschaft
und Linguistik
Postfach 10 01 31
D-33501 Bielefeld
Germany
brose@nov1.lili.Uni-Bielefeld.de
tel.: +49-521-106 53 07

Gerhard Strube
Universität Freiburg
Institut für Informatik und
Gesellschaft /Kognitionswissenschaft
Friedrichstr. 50
D-79098 Freiburg
Germany
strube@cognition.iig.uni-freiburg.de
tel.: +49 761 203 49 33/34

Patrizia Violi
University of Bologna
I.D.C.
Via Selvatica 2
I-40123 Bologna
Italy
mo7bova1@icineca.bitnet
tel.: +39-51-58 15 34

Wolfgang Wahlster
DFKI Saarbrücken
Stuhlsatzenhausweg 3
D-66123 Saarbrücken
Germany
wahlster@dfki.uni-sb.de
tel.: +49-681-302 5252

Ralph Weischedel
BBN Systems and Technologies
70 Fawcett Street
Cambridge MA 02138
USA
weischedel@bbn.com
tel.: +1-617-873-3496

Alexander Sigel
FH Hannover
Fachbereich IK
Hanomagstr. 8
D-30453 Hannover
Germany
sigel@www.bid.fh-hannover.de
tel.: +49 (511) 44 43 44

Zuletzt erschienene und geplante Titel:

Ph. Flajolet, R. Kemp, H. Prodinger (editors):
"Average-Case"-Analysis of Algorithms, Dagstuhl-Seminar-Report; 68; 12.07.-16.07.93 (9328)

J.W. Gray, A.M. Pitts, K. Sieber (editors):
Interactions between Category Theory and Computer Science, Dagstuhl-Seminar-Report; 69; 19.07.-23.07.93 (9329)

D. Gabbay, H.-J. Ohlbach (editors):
Automated Practical Reasoning and Argumentation, Dagstuhl-Seminar-Report; 70; 23.08.-27.08.93 (9334)

A. Danthine , W. Effelsberg, O. Spaniol, (editors):
Architecture and Protocols for High-Speed Networks, Dagstuhl-Seminar-Report; 71; 30.08.-03.09.93 (9335)

R. Cole, E. W. Mayr, F. Meyer a.d.Heide (editors):
Parallel and Distributed Algorithms, Dagstuhl-Seminar-Report; 72; 13.09.-17.09.93 (9337)

V. Marek, A. Nerode, P.H. Schmitt (editors):
Non-Classical Logics in Computer Science, Dagstuhl-Seminar-Report; 73; 20.-24.09.93 (9338)

A. M. Odlyzko, C. P. Schnorr, A. Shamir (editors):
Cryptography, Dagstuhl-Seminar-Report; 74; 27.09.-01.10.93 (9339)

J. Angeles, G. Hommel, P. Kovács (editors):
Computational Kinematics, Dagstuhl-Seminar-Report; 75; 11.10.-15.10.93 (9341)

T. Lengauer, M. Sarrafzadeh, D. Wagner (editors):
Combinatorial Methods for Integrated Circuit Design, Dagstuhl-Seminar-Report; 76; 18.10.-22.10.93 (9342)

S. Biundo, R. Waldinger (editors):
Deductive Approaches to Plan Generation and Plan Recognition, Dagstuhl-Seminar-Report; 77; 25.10.-29.10.93 (9343)

P. Gritzmann, D. Johnson, V. Klee, Ch. Meinel (editors):
Counting Issues: Theory and Application, Dagstuhl-Seminar-Report; 78; 06.12.-10.12.93 (9349)

B. Endres-Niggemeyer, J. Hobbs, K. Sparck Jones (editors):
Summarizing Text for Intelligent Communication, Dagstuhl-Seminar-Report; 79; 13.12.-17.12.93 (9350)

Ch. Brink, G. Schmidt (editors):
Relational Methods in Computer Science, Dagstuhl-Seminar-Report; 80; 17.01.-21.01.94 (9403)

A. Arnold, H. Seidl, B. Steffen (editors):
Algorithms in Automata Theory, Dagstuhl-Seminar-Report; 81; 07.02.-11.02.94 (9406)

K. Ambos-Spies, S. Homer, U. Schöning (editors):
Structure and Complexity, Dagstuhl-Seminar-Report; 82; 14.02.-18.02.94 (9407)

H. Noltemeier, T. Ottmann, D. Wood (editors):
Data Structures , Dagstuhl-Seminar-Report; 83; 28.02.-04.03.94 (9409)

J. P. Finance, S. Jähnichen, J. Loeckx, D. Smith, M. Wirsing (editors):
Logical Theory for Program Construction, Dagstuhl-Seminar-Report; 84; 07.03.-11.03.94 (9410)

R. Klette, W.G. Kropatsch, F. Solina (editors):
Theoretical Foundations of Computer Vision, Dagstuhl-Seminar-Report; 85; 14.-18.03.94 (9411)

A. Buchmann, S. Chakravarthy, K. Dittrich (editors):
Active Databases, Dagstuhl-Seminar-Report; 86; 21.03.25.03.94 (9412)

F. Meyer a.d. Heide, H.J. Prömel, E. Upfal (editors):
Expander Graphs, Random Graphs and Their Application in Computer Science, Dagstuhl-Seminar-Report; 87; 11.04.-15.04.94 (9415)

J. van Leeuwen, K. Mehlhorn, T. Reps (editors):
Incremental Computation and Dynamic Algorithms, Dagstuhl-Seminar-Report; 88; 02.05.-06.05.94 (9418)

R. Giegerich, J. Hughes (editors):
Functional Programming in the Real World, Dagstuhl-Seminar-Report; 89; 16.05.-20.05.94 (9420)

H. Hagen, H. Müller, G.M. Nielson (editors):
Scientific Visualization , Dagstuhl-Seminar-Report; 90; 23.05.-27.05.94 (9421)

T. Dietterich, W. Maass, H.U. Simon, M. Warmuth (editors):
Theory and Praxis of Machine Learning, Dagstuhl-Seminar-Report; 91; 27.06.-01.07.94 (9426)

J. Encarnação, J. Foley, R.G. Herrtwich (editors):
Fundamentals and Perspectives of Multimedia Systems, Dagstuhl-Seminar-Report; 92; 04.07.-08.07.94 (9427)

W. Hoeppner, H. Horacek, J. Moore (editors):
Prinzipien der Generierung natürlicher Sprache, Dagstuhl-Seminar-Report; 93; 25.07.-29.07.94 (9430)

A. Lesgold, F. Schmalhofer (editors):
Expert- and Tutoring-Systems as Media for Embodying and Sharing Knowledge, Dagstuhl-Seminar-Report; 94; 01.08.-05.08.94 (9431)

H.-D. Ehrich, G. Engels, J. Paredaens, P. Wegner (editors):
Fundamentals of Object-Oriented Languages, Systems, and Methods, Dagstuhl-Seminar-Report; 95; 22.08.-26.08.94 (9434)

K. Birman, F. Cristian, F. Mattern, A. Schiper (editors):
Unifying Theory and Practice in Distributed Systems, Dagstuhl-Seminar-Report; 96; 05.09.-09.09.94 (9436)

R. Keil-Slawik, I. Wagner (editors):
Interdisciplinary Foundations of Systems Design and Evaluation, Dagstuhl-Seminar-Report; 97; 19.09.-23.09.94 (9438)

M. Broy, L. Lamport (editors):
Specification and Refinement of Reactive Systems - A Case Study, Dagstuhl-Seminar-Report; 98; 26.09.-30.09.94 (9439)

M. Jarke, P. Loucopoulos, J. Mylopoulos, A. Sutcliffe (editors):
System Requirements: Analysis, Management, and Exploitation, Dagstuhl-Seminar-Report; 99; 04.10.-07.10.94 (9440)

J. Buchmann, H. Niederreiter, A.M. Odlyzko, H.G. Zimmer (editors):
Algorithms and Number Theory, Dagstuhl-Seminar-Report; 100; 10.10.-14.10.94 (9441)

S. Heinrich, J. Traub, H. Wozniakowski (editors):
Algorithms and Complexity for Continuous Problems, Dagstuhl-Seminar-Report; 101; 17.10.-21.10.94 (9442)

H. Bunke, T. Kanade, H. Noltemeier (editors):
Environment Modelling and Motion Planning for Autonomous Robots, Dagstuhl-Seminar-Report; 102; 24.10.-28.10.94 (9443)

W. Maass, Ch. v.d. Malsburg, E. Sontag, I. Wegener (editors):
Neural Computing, Dagstuhl-Seminar-Report; 103; 07.11.-11.11.94 (9445)

G. Berry, W.P. de Roever, A. Poigné, A. Pnueli (editors):
Syncronous Languages, Dagstuhl-Seminar-Report; 104; 28.11.-02.12.94 (9448)