Summary writing strategies based on discourse structures, relevance theory and kernel preservation

Michael P. Jordan
Queen's University at Kingston

This paper explores methods of guiding summary writers based on the discourse structure of the material being summarized. Multi-item discourse structures are identified, and the problem-solution macrostructures (of special interest for technical writing) are used as an example of how we can preserve the "gist" (content plus organization) of the original message in such genres. A similar approach is taken for the established binary logical relations of discourse connection, which can form the basis of text macrostructure or microstructure.

Descriptive texts do not have such macrostructures on which to base summaries. For these, the advice provided consists in selecting high-priority items of information based on adaptations to relevance theory. Most texts are recognized as having many structural levels, with different types of structure at each level. The approach adopted for summarizing such texts is progressive summarization of the text strata using the methods appropriate for the structure at each level.

The need to identify and retain the central "kernel" or "essence" of the original material, as well as the gist, is explained and demonstrated. This highlights the need for skilled judgment in selecting vital material for inclusion in the summary – something computer summarizing tools cannot yet accomplish. Deletion heuristics for summarizing are provided based on established text structures.

Dans cet article, on examine diverses façons de rédiger des résumés en s'appuyant sur la structure même du texte de départ. D'abord, on expose diverses structures discursives à composantes multiples. On démontre comment la structure explicative, servant à proposer des solutions à un problème et qui dessert bien la rédaction professionnelle, permet de cerner la thématique du texte de départ soit l'ensemble structuré des idées essentielles du texte. On souligne que les structures binaires ont cette même propriété et peuvent constituer soit la macrostructure d'un texte, soit sa microstructure.

Les textes descriptifs, quant à eux, ne possèdent pas de telles macrostructures sur lesquelles s'appuyer pour rédiger des résumés. Aussi, on propose de relever les points saillants de ces textes au moyen de la théorie de la pertinence. En général, les textes sont structurés en niveaux et comportent des structures différentes à chacun de ces niveaux. Pour résumer de tels textes, il convient d'utiliser une démarche propre à la structure de chacun des niveaux.

De plus, on explique comment relever l'idée dominante ou l'essence du texte de départ (kernel), de même que sa thématique (gist). Cette démarche démontre qu'on doit faire appel à un jugement éclairé pour choisir les données à inclure dans le résumé, jugement dont ne peut se prévaloir la réduction assistée par ordinateur. On propose, enfin, des procédés heuristiques pour supprimer les éléments non-pertinents à la rédaction de résumés en s'appuyant sur des structures de texte établies.

Background

Different Approaches

Russell's (1992) exploration of methods of describing and classifying summaries encompasses structural, metatextual, cognitive and contextual criteria. The metatextual function deals with overt reference in the summary to elements of the larger document being summarized, the overriding criterion for descriptive abstracts (Jordan, 1991a). The so-called cognitive approach actually stems from structural considerations. It is based on Kintsch and van Dijk's (1978) processing model of discourse comprehension, in which the summarizer employs a number of rules (such as deletion, generalization and construction) to reconstruct, in miniature form, the macrostruc-

ture of the original document. Contextual factors include information about the communicative source of the original text (Laurent, 1985, p. 84); such summaries may embrace wider considerations than the document and its summary, also delving into the overall situation of the document and its summary, the purpose of the summary as a communicative act, and the roles of those engaged in the writing. This echoes the "field," "mode" and "tenor" of systemic linguistics (e.g., Halliday, 1978, p. 142–143). Recent analysis (Jordan, 1999b) of a "popularized summary" (Russell, 1994, p. 38) in science shows how the summary writer often goes well beyond the contents of the original article in explaining the significance and relevance of the original work.

Summaries have also been studied from the perspective of teachers seeking to understand and improve their students' cognitive and linguistic skills, from the point of view of information scientists concerned primarily with information retrieval, and from the viewpoint of linguists describing and classifying types of document structure. Those involved with genre analysis, the cognitive science of document comprehension, and artificial intelligence and computer "autosummarizing" tools also have a strong interest in summaries—as of course do teachers of technical and business communications.

This paper provides a critical summary and analysis of the use of various types and levels of information structure as the basis for creating useful summaries, encompassing both the "structural" and the "cognitive" techniques identified by Russell (1992). The work will encompass summaries that condense the total body of information available on a specific topic ("summary documents"), as well as those that summarize one specific document, which may or may not accompany the summary ("document summaries"). The latter, however, will be restricted to summaries that do not extend the information beyond that which is included in the document being summarized. That is, the important area of study dealing with the use of a document summary as part of a wider communication function is not included here.

The perspective adopted is that of a practical teacher of technical writing seeking to understand and explain the structural foundations on which summaries can be created and critiqued. As the interests of those seeking to find objective criteria for the automatic creation of summaries overlap with the main interests here, they are discussed when they become relevant. The interests of cognitive scientists, especially with respect to relevance theory, also overlap the main interests here, and these are also discussed when appropriate.

The Structural Approach

Beaudet (1994, p. 52–58) uses structural models of description, narrative, exposition, argument and instruction not just as the basis for student comprehension of the subject to be summarized, but also as an aid to the summarization process itself. This approach is reflected in a number of French language books on the subject (Boret and Peyrot, 1971; Moreau, 1981; Valentine, 1990). The structures of informative abstracts themselves have been shown, in general terms, to follow that of the main document: scope, purpose, methods, results, conclusions (Cremmins, 1982; Collison, 1971).

Arguably, all well-organized texts are "structured" in the sense that they follow some sort of definable and often predictable pattern of information. Yet some are more structured than others, following well-established sequences of information and using well-known signals of transition between the types of information presented. Some of these texts may follow narrative sequences, while others follow the "Problem-Solution" pattern or its many variants; still others follow the simpler binary logical pairs of Cause-Effect, Purpose-Means, Assessment-Basis, and so on. The types and patterns of information in texts that are less structured in these ways may still be understood in terms of the Five Ws of journalism, or as descriptions in which certain types of information (and their order) are typical, though perhaps not statistically predictable. In practice, of course, most texts have structures of different types at separate levels, or strata, of the communication. We examine here how our understanding of these structures can form the basis for selecting the contents and order for suitable summaries.

After an analysis of some basis principles of summary creation, this paper explains the use of the three major branches of text organization: the multi-item structures of narrative and Problem-Solution patterns, the logical and general binary relations of meaning, and the more general or *ad hoc* patterns of descriptive text. The separate treatment of each type allows us to recognize practical strategies for writing summaries for texts of each discourse pattern. Then, by elaborating on van Dijk's (1977) notion of macrostructure (or "superstructure" in van Dijk and Kintsch, 1983) and microstructure in text, we can developed summary writing strategies for complex texts which exhibit several types of structure at different levels of their development.

Some General Principles

Importance and Gist

In general, we need to recognize that the summary should "contain all the necessary information and nothing but this information.... The major problem is, of course, how to determine which parts of a text are of prime importance and must be reproduced in the shorter version" (Fries, 1987, p. 48). This problem is tackled here by first recognizing the macrostructure of the information to be summarized and then identifying the high-priority types of information within that structure, i.e., the "important" information to be included, in brief form, in the summary. For descriptive texts, the structural guidance may be much weaker, and we then have to rely more on the reader's needs and the purpose of the document as selection criteria; this will involve the need for tighter definitions for the terms "necessary" and "importance" used above by Fries. For texts of mixed structure, both techniques are used.

Fries's view about important information is supported by van Dijk, who notes that the summarizer needs to "answer what the story is about" based on the concept that "the topic is determined by what, from some perspective, seems the most important fact(s) of the story" (1981, p. 187). The question as to what is most "important" is raised later in relation to all forms of structure discussed here, and it is vital in deriving criteria for general, descriptive texts. But generally we can take it as meaning what is most relevant to the needs of readers and the purpose of the communication. For general stories, it might mean "most interesting" or "most entertaining"; for educational texts, it might mean "most informative"; for hortatory texts, it might mean "most useful"; for persuasive texts, it might mean "most convincing" or "most stimulating," etc. That is, the meaning of "importance," noted by both Fries and van Dijk, is recognized here as a purpose-specific concept.

Many writers express the importance of material to be summarized in terms of the "gist" of that material, e.g.:

> He [the summarizer] is then ready for a second more careful reading
> of the author abstract (if any), the first and last paragraphs, and key
> sections in the original document with headings such as "Introduc-
> tion," "Purpose," "Conclusions," "Summary," and "Recommendations."
> These paragraphs contain the *gist* of what the author considers to be
> important and hence are important to the abstractor in identifying
> document contents. (Maizell *et al.*, 1971, p. 77)

Brown and Day (1983, p. 2) discuss the importance of the summarizer's inclusion of the gist in one's own words, and Rino and Scott (1996) present a discourse model for "gist preservation." Bower and Cirilo (1985) base their cognitive approach to summarizing and memory retention on Kintsch and van Dijk's (1978) model of text comprehension, which "describes the global organization of a text, its *gist*, thus making summarization and long-term memory of the ideas in the text manageable tasks" (p. 92).

Text comprehension is also the basis for Baker *et al.*'s (1988, p. 65-66) use of Pearson and Camperell's definition of gist: "[E]lectrical engineers formulated *gist* units to represent the macrostructure of the text and information units to represent its microstructure...Each *gist* unit consisted of a one sentence summary of the main idea of the text segment (1981, p. 57). Phillips also relates gist to memory recall by noting that "It is the 'gist' of a text that is recalled rather than its wording" (1985, p. 4).

Gist, Essence, and Kernel

These definitions of gist are clearer than is van Dijk and Kintsch's 1983 definition: "Such a macrostructure is the theoretical account of what we usually call the gist, the upshot, the theme, or the topic of a text" (1983, p. 15). This first states that gist is the macrostructure of the text ("that which defines its global coherence" (van Dijk, 1983, p. 115)) and then equates it to theme or topic. This confusion is not repeated later: "[F]rom this textbase, the macrostructure is derived representing its essence or gist" (Kintsch, 1985, p. 231).

Pechenik also uses the term "essence" to convey the important types of information that need to be included in the summary (or abstract): "It [the abstract] must completely summarize the *essence* of your report: why the experiment was undertaken; what problem was addressed; how the problem was approached; what major results were found; what major conclusions were drawn" (1993, p. 104). The meaning of "gist" and "essence" is generally assumed to be the total amount of important information to be included in the summary together with the text macrostructure that that represents, i.e., Rino's (1996, p. 27) "*esqueleto básico*" (basic skeleton) of the text. Van Dijk's use of "the most important fact(s) of the story" and Baker *et al.*'s use of gist as a central feature of a text segment, however, do recognize that the gist or essence of the story (or segment of a story) can be a single item of information, whereas the others cited above clearly intend it to mean a structured sequence of different types of information.

That the essence is the central feature of a story is apparent in Rino's (1996) use of *Idéia Central* in the title of her major work. Her term *esqueleto básico* encapsulates the "content + structure" meaning of "gist," but she defines this as a *proposição central*

de um discurso (p. 27), defined in Rino and Scott (1996) as: "The central proposition. This is the 'kernel' of the discourse... i.e., the information around which the discourse is organized to satisfy the communicative goal" (p. 2). Although they admit this central proposition can be highly complex, they restrict it, with reasonable validity, to a single component.

Both types of central or important information are assumed here: the "gist," which comprises all the individual items of information central to an acceptable discourse on the topic together with the macrostructure that that information creates; and the "kernel," which is the single (occasionally more) feature of information without which the story makes no sense, or is of no use, or has no news value—depending on the document's purpose. The difference is the difference between Kintsch's "gist as macrostructure" and Rino's *Idéia Central.* The distinction between the two—and the importance of the kernel of an account for summary writing—are explained and demonstrated later.

Document Summaries and Summary Documents

The broad view taken here is that the summary is a short document representing a larger amount of information. For the "document summary," that source of information is the original document, which may or may not accompany the summary; but for the "summary document," the source is the total pool of information available about the topic. In both cases, the summary presents the most essential parts of the information available. This means we need not be concerned in our definition of "summary" whether the larger amount of information has actually been expressed as a text or document; the summary can summarize an accompanying text, a separately published document, or material as yet unarticulated as a document.

More far-reaching, perhaps, is the contention that the total information available about a subject is already "structured" whether or not it has yet been articulated in some sort of text. As examples, the five Ws of a news story exist even before the story is written, details of an aircraft accident are connected as an Effect-Cause set of information even though much of it is yet to be discovered, and the creation of a solar-powered bicycle is basically a Problem-Solution set of information even though parts of it have yet to be designed. Thus structure is viewed here as the ways different types of *information* cohere together and are reflected in the document that summarizes that information, leaning on Winter's (1976) work on the *Fundamentals of Information Structure.*

This approach, in line with my earlier claim (1984a, p. 9) that "any description is a *summary* of the total information available for what is being described," allows us to recognize the writer's role in selecting information for any new document as essen-

tially summarizing from the total information on the subject being described. It also allows us to recognize the occasional need for summaries to go beyond the text information itself to include the wider contextual considerations discussed earlier. In this way, too, we can recognize popularized texts like the one discussed in Jordan (1999b) as having two functions: they summarize another document, but they also include additional material (using a broader informational structure) which summarizes some of the wider contextual information available about the subject matter.

Conciseness and Construction

Russell (1979, p. 19) discusses the distinction between conciseness and summarizing. She notes that "When you see him at a social event, you have to wonder if he were born in a barn" can be summarized by "He is socially inept" as a more concise form. Here, conciseness is viewed in a more technical way, as creating a shorter form that says much the same as the original. This can be distinguished from the summary in Russell's examples above, which share a Particular-General relation. As shown later, such relations form the basis for summarizing by replacing "the given level of linguistic detail and abstraction by a cotextually higher level of abstraction" (Werlich, 1976, p. 86). Although the distinction between conciseness and summarizing may become blurred in some instances, the difference in principle is important: conciseness is the same information in a briefer form, whereas a summary presents less information and/or more-general information.

In Russell's summary above, we see that the words of the summary do not occur in the original text. Van Dijk (1981, p. 178) makes this point by noting that: "[A] summary is based on a construct, 'taking together' semantic information from the discourse as a whole." This view is supported by Hidi (1983, p. 4), who advocates the recombination of ideas "into novel configurations." Sherrard (1989, p. 1) expresses this in terms of the "deep level transformation of a text, departing from surface wording and original order of the propositions." And Brown and Day (1983, p. 2) suggest that rearrangement of material and statement of the gist in the summarizer's own words are marks of a mature summarizer's work. Winograd (1984), Johns (1985) and Sherrard (1989) go further by noting that the ability to combine and integrate ideas across paragraph boundaries is a criterion for mature summary creation. See Russell (1994, p. 43) for a more detailed presentation.

Using one's own words in a summary may be desirable in some situations, but there are exceptions: brief evaluations and recommendations, for example, may be better rendered by verbatim repetition. In computer summarizing (called *autosummarizing*) however, such an approach could create unrealistic goals as the extraction of existing paragraphs, sentences and phrases is much easier for the com-

puter than the construction of new expressions. As a result, summary-creating heuristics depend quite noticeably on the aims of the researchers—especially on whether they are developing strategies for human or for machine use.

Reduction and Deletion

Both van Dijk (1977) and Manning (1990) claim that a text can share the same semantic structure as an individual sentence. Mann and Thompson (1987, p. 40) make a stronger claim about the inter-strata applicability of units of information: that "the same sorts of relations characterize text structures at all levels." Jordan (1998a) goes even further with his analysis of the Cause-Effect relation, showing that this relation occurs not just between spans, paragraphs, sentences and clauses, but also within the clause—and even within the noun phrase! This allows us to summarize Cause-Effect texts at the briefest levels of language structure, including headings and titles. As an example, two levels of summary are included in the headline and first sentence of the following example (from Jordan, 1997, p. 334):

(1) James Mackie killed in accident
    A Kingston township man was killed Tuesday *when* his jeep skidded off an
    icy highway near Lindsay and slammed into a tree. (*Kingston Whig
    Standard*, February 29, 1996, p. 9) (italics added here and later)

The first sentence is a summary of the Effect-Cause structure of the remainder of the article, and the headline is an even briefer summary of the first sentence. The preposition "in" in the headline and the time adverbial "when" in the first sentence signal the logical relation; see Jordan (1997) for detailed discussion of such subtle logical signalling in news reporting.

It is possible to write such Effect-Cause summaries at different levels of detail because the *information* is essentially an Effect-Cause set. A person has been killed (Effect) and something (the Cause) has caused this to happen. We can elaborate on or reduce both elements of this set, but we are still left with the same structure whether we express this as a simple sentence, two sentences, two paragraphs, two chapters, etc. Manning (1990, p. 374) uses this principle when he notes that: "We summarize a unit of discourse by substituting a smaller semantically equivalent unit," using "extended text," "paragraph," "argument," "sentence" and "phrase" as his units for reduction. This view was expressed earlier by Werlich (1976, p. 86), who claimed such a reduction as a principal method of summarizing: "[The summarizer] replaces the given text units by cotextually lower-level text units, e.g., replaces a text of chapter length by a summary of paragraph length." Clearly, as shown above, we can use much smaller text units in applying this principle. In this way, we can create summaries by retaining

the semantic meanings and relations of the original information or the original docu-
ment, while reducing the length by saying less about each type of information. This
means a summary of an Effect-Cause will have an Effect-Cause structure, a summary
of a Problem-Solution text will have a Problem-Solution structure, etc.

An additional strategy, especially for microstructural levels of the original text, is
the deletion, rather than the reduction, of types of information. This approach is
noted by Russell (1979, p. 14) as follows: "In some structures, a primary idea might be
explicitly worded, surrounded by details like china packed in excelsior. In such cases,
the basic information can be easily extracted." Russell's simile is general enough to
apply to all types of texts and their summaries. For semantically-structured texts, we
can often recognize one or more elements of the account as being the wrapping (or
less important information) from which the more important information can be ex-
tracted. As an example, the headline in Example 1 excluded the *cause* of the acci-
dent—the icy road; although important enough for the lead sentence, this informa-
tion was regarded as not important enough to cram into the headline. This approach
is in line with the structure-based deletion heuristics developed by Rino and Scott
(1994, 1996) for computer summarization, which are discussed later. For the more
general case of descriptive writing, where the semantic boundaries and types of in-
formation are less clear, we will have to use other more subjective criteria.

Summarizing Multi-Item Structures

Multi-Item Structures

Many texts are structured around one or more of the recognized language-inde-
pendent semantic structures that organize material in accordance with traditional
types of information and sequences. This major section deals with structures that
contain more than two types of information, and the next section deals with the
binary systems, such as Effect-Cause. An early multi-item sequence is Burke's (1945/
1969, p. xv) "dramatistic pentad" of Act-Agent-Scene-Agency-Purpose based on analy-
sis of stage events. Both Beardsley (1950) and Young and Becker (1966) note the se-
quence of Topic-Restriction-Illustration, and Labov and Waletsky (1967) and Labov
(1972) identify the pattern of Abstract-Orientation-Complication-Action-Evaluation-
Result/Resolution-Coda in oral narratives. Longacre (1972) identifies a complex pat-
tern for some narratives: Aperture-Setting-Inciting moment-Developing conflict-
Climax-Dénouement-Final suspense-Closure.

Of greater interest for summaries of technical writing, Winter (1976) identified
the Situation-Problem-Solution-Evaluation pattern for technical reports. Also van
Dijk (1977) notes both the narrative structure Setting-Complication-Resolution-Evalu-

ation-Moral and the Introduction-Problem-Solution-Conclusion pattern for scientific discourse, which is essentially the same as Winter's pattern. In analyzing abstracts for empirical scientific work, several workers (e.g., Cremmins, 1982; Jordan, 1991a) have noted the structure as Scope-Purpose-Method-Results-Conclusions. In creating summaries for such time- and activity-oriented structured texts — brief details for each of the major elements of the account can provide a satisfyingly complete report.

The use of the macrostructure of the original material as the basis for the summary is noted by Katz (1985, p. 70) as including not just the themes, or major types of information, but also the "logic" or structure of the original material: "The abstract should be a guide to the reader by pointing out major themes and foreshadowing the logic: it is a portable microcosm of your work." The idea of the summary as a microcosm of the macrostructure (in both content and structure) of the original material is a useful one, which is explained in more detail for Problem-Solution texts below.

Problem-Solution Structures

The problem-solution macrostructure is of great importance in engineering writing because engineers create solutions to problems in and of society, and the documents that explain such activities and products naturally follow the problem/solution procedure of those engaged in the work. Science is different in that the "problem" for scientists is the intellectual need to know or understand what something is or how or why something works, and the "solution" is the answer to this question, in terms of an explanation, model, formula, etc. In business and commerce too, we constantly need to solve problems, and thus many texts in these fields also follow Problem-Solution structures. (See Jordan, 1984b, for an analysis of over 100 examples of such texts from all walks of life.) Because of its prevalence in, and importance to, technical and business writing, the Problem-Solution structure is used here as an example of how such structures can be used to help us write effective summaries for multi-item structures.

Both Beardsley (1950) and Young and Becker (1966) mention problems and solutions in terms of the wider structures of information. Hutchins (1977a, 1977b) also discusses this informational pair in technical texts and with specific mention of its usefulness for abstracting purposes. He views Problem-Solution as one of a series of "oppositions," based on a pattern of expectation first discussed by Sweet (1891) with his "Try-Succeed" pair. In 1976, Winter produced a pilot manual placing the central Problem-Solution pair within a wider framework of "Situation-Problem-Solution-Evaluation," which formed the basis for later work by Hoey (1979, 1983) and Jordan (1980, 1981, 1984b). This macrostructure is also identified by van Dijk (1977) as the "scientific discourse structure" Introduction-Problem-Solution-Conclusion.

The use of Winter's and van Dijk's four-part structure differs from the approach taken by Rino and Scott (1996) and the central premise of Mann and Thompson's Rhetorical Structure Theory in that it does not recognize a single central "proposition" for the document; rather it relies on the inclusion of all relevant parts of the structure for completeness of the discussion. Using a basic four-part made-up "story" of "I was on sentry duty. I saw the enemy approaching. I opened fire. I beat off the enemy.", Winter noted that this constitutes a minimally-complete account. Removal of any of the four components, he claims, would leave the account incomplete. (See Hoey 1983, p. 31–61 for a very detailed discussion of this example.) For such a text, there is no single central proposition; instead all parts together form the essential information that must be included in a meaningful summary. Put another way, the "gist" for such sets of information is a composite of all four parts of the overall structure. As we saw earlier, this term is also used by others (e.g., Brown and Day, 1983, p. 2; Basham, 1986, p. 110) to refer to several parts of a document rather than just its central item, or kernel, of information.

The Four Parts as Summary

Hutchins (1977a) was aware that the four parts of such structures could form a basis for an adequate summary or "abstract." In fact he argues against such a procedure on the grounds that, although it may produce an admirable précis, it leaves out too much of the essential material to make up an good abstract. Hoey (1983, p. 80), however, notes that use of the four-part structure does provide a satisfactory skeleton summary, reasoning that it "gets to the communicative core of the discourse." He observes: "a reasonable skeleton summary of the discourse can be achieved by the simple expedient of taking the first sentence of each element [Situation, Problem, Solution, Evaluation] of the pattern."

By including small parts of each of the four parts in a summary of a text, Jordan (1984b, p. 15, 27–28,) shows how we can use our knowledge of the components of the structure to create good summaries for this genre. His approach differs from that advocated by Hoey in that it uses selected important *information* in each of the four parts, rather than mechanistically taking the first sentence of each part. Hoey's method may be more suitable for autosummarizing tools, although Hoey notes the need to exclude some of the signalling words, a task the computer may find difficult. But it is still not always likely to yield the best summary. For that, we usually need to create a *construct* (van Dijk, 1981, p. 178) using our own words from each of the four parts of the macrostructure (see earlier discussion).

Multiple Levels of Summary

The argument goes further. In his work on very short texts (1981) and his chapter "Short Texts as Summaries," Jordan (1984b p. 8–19) shows how four-part Problem-Solution structures (however small or large) are themselves summaries of the total information available on the subject. He cites an article published in *Engineering Digest* which is overtly announced as a summary of a previously published article in *Design News* to demonstrate two levels of summary: original material → first article in *Design News* → summary in *Engineering Digest* of the first article. By expanding the well-known two-dimensional "Pyramid" Technique (e.g., Blicq, 1983, p. 314–317) with the summary at the apex of the triangle, he shows different levels of information selection for different types of information within the same document. This explains the occurrence of all four information categories in the summary at the start of the second article (already a summary of a summary):

   (2) It now appears from the results of recent work carried out by the Metal
        Improvement Co. of Teaneck, NJ, that high-intensity peening of 300 Series
        austenitic steels will also prevent intergranular corrosion cracking. (first
        sentence of an article in *Engineering Digest*, April 1976, p. 16, summarized
        from *Design News*, January, 1976, p. 12)

This summary of a summary of a summary of the total information available contains all four parts of the Problem-Solution macrostructure: the Situation dealing with the company, the Problem of intergranular corrosion cracking, the Solution of high-intensity peening, and the Evaluation (*will prevent*) that the shot peening is indeed a solution to the problem.

The very effective title for this article also contains these four components of the total details available:

   (3) Peening Process Prevents Intergranular Corrosion of Stainless Steels *(ibid.)*

Again the four parts of the story are included to provide a minimally complete account as a very short summary—a summary of a summary of summary of a summary of the total information available! The effectiveness of such summary-titles relies on its containing information about each of the four parts of the macrostructure. In terms of Kintsch and van Dijk's (1978, p. 376) notion of a summary being a "second-order discourse" about the original document, this title is a "fourth-order discourse."

Such summary titles occur in the active voice, as shown above, or the passive voice, as in "Costs at New Brunswick Coal Cut by an On-Line Monitoring System" (Jordan, 2000a, p. 107), which includes the Situation (at New Brunswick Coal), the Problem ([high] Costs), the Solution (an On-Line Monitoring System) and the Evaluation (the verb cut). These examples demonstrate that, for such multi-item texts, we need to include in the "gist" information about each of the four parts to create a sound summary.

Using Parts of the Macrostructure

The overall theory of problem-solution texts (e.g., Jordan 1984b) includes recognition of texts that do not contain all four parts of the macrostructure. First, we need to recognize that, for many engineering documents, the four parts really represent an extremely involved sequence of thought and events—from a general situation, to a recognition and definition of a problem, through creation and refinement of solutions and the decision to proceed with the best one, and then to its completion, implementation and testing. There are actually many steps in this "design process" (e.g., Earle, 1977). However, progress reports will contain only those elements completed at the time of writing, and reports of abandoned projects will also include only the stages of work completed before abandonment. In addition, readers may know some of the information already. For any of these reasons, the document may not include all four parts of the macrostructure (Jordan, 1988, p. 13). Obviously, the summaries for such documents will include only the information available or appropriate for the audience and may not include all four parts.

Based on his earlier (1980) 12-part algorithm of the problem-solution sequence, Jordan (1991a) explains that many summaries in informal science texts are examples of the more general Problem-Solution framework. The titles and/or initial summaries of many of the texts in his corpus (from *New Scientist*) concentrate on the problem, whereas others concentrate on the solution to a problem; still others have three or even four parts to represent successful completion of a problem-solving task. The discussion also includes interactive (also called "interpersonal" or "people") problems, in which the solution planned or implemented by one person or group is perceived as a problem by or to another person or group. Such complications are seen in the following example:

(4) Animal Researchers to get Legal Aid
    Legal aid will soon be available for scientists, doctors and vets who seek
    redress when they feel they have been libeled by anti-vivisection groups.
    (*New Scientist*, February 17, 1990, p. 20)

This opening summarizes the larger document, which explains vivisection as both a solution to the need for medical research and a perceived problem to anti-vivisectionists. Their solution in trying to have the practice stopped is to make uncomplimentary remarks about those engaged in the practice. This, in turn, is viewed as libel by (a problem for) the scientists, doctors and vets involved, and their planned legal solution (which will be a problem for the anti-vivisectionists) requires financial aid as a solution to their funding problem in seeking to implement the solution. The legal aid is a solution to the financial need. This very complex set of interpersonal perceptions and actions is well summarized in the title and initial sentence given in the text.

In more formal scientific documents, Problem-Solution structures are still found to occur in summaries, but the other logical relations and other linguistic systems (e.g., "known-new" and other "contrast" pairs) are also used (Jordan 1991c). For very formal documents describing projects in pure and applied science, the summaries become more complex (Jordan, 1993). Problem-Solution components are again still present, but the problem is usually more of the "need-to-know" type, in which investigation leads to an understanding or an explanatory (often mathematical) model of the behaviour investigated. For the early parts of many technical papers, the texts and their summaries often follow the general research investigation pattern of "Topic definition and description–Importance of topic–Related work– Precise task (problem)– Methods used" (Jordan, 1993). Such a pattern is shown to be the basis for combined Introduction/ Summary openings for large formal technical papers (Jordan, 1991a).

Other Multi-Item Structures

Although emphasis here has been on the four-part problem-solution structure, that should be seen as just one example of how multi-example structures can be used as the basis for creating useful summaries. Another useful structure is Statement-Denial-Correction (Hoey, 1983, p. 128–129), in which a statement is given and then denied (stated as being incorrect), followed by a "correction" (a statement of what is true). The related structure of Statement-Denial-Basis for denial ends with justification for the denial—and of course we can have both Correction and Basis to create a four-part structure. Contrasted sentences, although often treated as binary systems, are really three-part structures in their full forms including the implicit comparative denial (something is true for X; it is not true for Y; what *is* true for Y (correction)). Similarly, concessive relations, although again traditionally treated in binary terms, involve four items in their full form (Statement-Expected conclusion-Denial of expected conclusion-Correction), and a fifth element of Basis for correction could also be added. A related three-part relation is Thesis-Concession-Rebuttal, in which the writer first makes a statement, then concedes a point apparently contrary to that the-

sis, and finally explains why that point does not in fact invalidate the original thesis (Werlich, 1976, p. 260). All these structures, which have denial as their central feature (Jordan 1998b, p. 724-729), can form the structural basis for creating summaries.

The broader structures mentioned earlier also form the basis for good summaries. The summary of a play, ballet or opera, for example, will probably follow Burke's dramatistic pentad, while historical accounts and developments are more likely to follow established narrative schemas. As shown for the Problem-Solution structure, all these other structures can form the basis for the inclusion in the summary of the high-priority parts of the total information available to the summarizer. And once the framework of the summary has been established in this way, less-important elements can be deleted and/or reduced to create a summary of appropriate length.

The Five Ws

Perhaps echoing Burke's pentad, the well-known journalistic formula or "Five Ws" (*Who?, What? When? Where? Why?*—plus *How?*) (e.g., Corbett, 1977, p. 27-29) is a somewhat different type of multi-item pattern, as it provides a useful guide to contents though not so reliably to structure. The *What?* of a story is arguably fundamental, although the extent of discussion for this item is highly dependent on the circumstances. Journalists are usually trained to emphasize the *Who?* element to introduce the "human element," which is usually claimed to increase interest in the story, although in technical writing of course this is largely or totally irrelevant. The other items of the account also depend very significantly on the situation. Thus, although the Five Ws (plus *How?*) may be a useful general guide for journalists, they need to be adapted to make them more useful for technical summary writers.

Some technical reports can be based loosely on the Five Ws principle, especially those involved with accident investigations. The writer needs to include in the report what happened, who was involved, why and how the accident happened (cause), and when and where. But the useful engineering element of such report (its kernel) is not a simple reporting of the facts (as the above categories might imply); rather it is the engineer's judgment regarding the contributory reasons for the accident, whether procedures or codes have been followed, and what procedures, components or people are at fault in creating the circumstances that led to the accident. Then meaningful conclusions can be drafted based on these assessments, and the all-important recommendations can be made. The summary for such reports should very briefly include information dealing with the Five Ws, but should concentrate on causes and related recommendations.

Summarizing Binary-Structured Texts

Origins and Relations

The main concern of those seeking to devise such multi-item patterns was to identify models of different genres of larger texts, although Burke (1969) was also concerned with the binary "ratios" between any two of his five major elements. Such binary relations formed the basis for early biblical translation work (Beekman and Callow, 1974), where information of a certain type is translated into the same type of information in the second language—and where the relationship between types of information also needs to be preserved. Binary semantic relations, or "semantic propositions," also enabled field linguists to comprehend the many languages in New Guinea (Longacre, 1972). From these studies and work by Hoey (1983), Jordan (1984b), Hobbs (1985), Mann and Thompson (1989) and others, we now have a sound understanding of the binary relations that deal not just with Problem-Solution, but also with Cause-Effect, Purpose-Means, Assessment-Basis, Enabler-Enablement, etc. These, as well as the more general relations of General-Particular and Generalization-Example (Hoey, 1983, p. 134-167), Part-Whole and Abstract-Instance (Hovy, 1990) and Unspecific-Specific (Winter, 1992), form a sound basis for creating useful summaries.

Although the Problem-Solution macrostructure is a powerful basis for constructing summaries, it is really a special organized pattern made up of several other simpler forms. For example, Jordan (1980) notes that evaluation occurs at many of the 12 stages in his Problem-Solution algorithm, and Hoey (1983) also discusses both Cause-Consequence and Basis-Assessment as components of the overall Problem-Solution framework. In addition, the Problem-Solution pair is very close to that of Purpose-Means: the Purpose is the decision to do something (often solve a Problem), and the Means is the way of doing it, i.e., the Solution. Engineers and scientists are also vitally concerned with a Cause-Effect relation in determining the cause of a problem before trying to overcome or solve it. Other multi-part structures also contain binary relations as essential components. As an example, one of Burke's (1969) binary "ratios" (the agency-purpose ratio) is the important Means-Purpose relation; also, as we have just seen, the three-part Statement-Denial-Basis for denial structure contains a binary Assessment-Basis relation as the denial is an assessment. This all means that, although we will initially treat the binary relations as isolated text structures of meaning, we must also recognize that they often form essential microstructure parts of more complex structures of informational and textual cohesion.

Cause-Effect and Assessment-Basis Relations

As we saw in Example 1, the total information about a topic and successive levels of summary about that information can be organized in a Cause-Effect manner—or Effect-Cause, of course. The ability to express this semantic connection at *all* levels of language—from the structure of a whole large document to within the clause, sentence or noun group—means that we can encapsulate the Cause-Effect meaning within summaries of two or three paragraphs, two or three sentences, or even within a simple sentence of a few words. The fact that this principle applies to all binary relations provides us with a powerful method of summary creation which faithfully reproduces the essential meanings and connection of meaning of the original information or document.

Information groupings, however, are rarely that pure or simplistic. Often the statement of a Cause-Effect relation is a conclusion or judgment (e.g., significant roughage in our diets reduces the incidence of colon cancer), for which Basis in the form of theoretical and/or empirical evidence may be necessary. Then the recommendations that may need to be made are assessments based on the conclusions reached. In science, the search for Cause-Effect relations (e.g., what causes ice storms) is essentially a problem to be solved, and we devise means or methods of determining the answers we seek; thus Problem-Solution and Purpose-Means relations are involved too. The need to understand what enables or inhibits actions or events also involves us with Enabler-Enablement relations (Jordan, 1998c) as well. Some of the complexities involved with such complex combinations are examined in Jordan (1999b).

The Assessment-Basis relation is not always neatly expressed in English—in fact many such pairs are not even contiguous in the text. We can often compress the relation originally expressed as paragraphs within a sentence using *based on* or *because*, *since* or *as* if a binary system of justification is adequate, as it often is (Jordan, 2001). However, when two or more premises are needed, or when two or more elements of basis need to be provided, or when the basis itself needs further basis to justify it, the logical explanation may require more than a subordinate clause-main clause sentence. It then becomes much more difficult to express the logical relations in a much briefer form than the original—especially as the omission of some parts of the reasoning could make the conclusion seem untenable. Deletion of the basis may be an option, as discussed shortly.

Purpose-Means and Problem-Solution

The closeness in meaning of the Problem-Solution and Purpose-Means pairs of information becomes extremely useful for writing summaries, a closeness shown by the Situation-Purpose-Means-Evaluation macrostructure (Jordan, 2000a, p. 104/105) that parallels Winter's (1976) Situation-Problem-Solution-Evaluation. The connection can be shown with the following example:

(5) To counteract the problem of accidental ignition of fluid in spray form when using hydraulic fluids, the British Standards Institute has produced a new draft for development, DD 61 'Flammability Spray Test for Hydraulic Fluids'. (*Safety*, April, 1979, p. 8)

The purpose *(To counteract the problem)* is the subordinate clause, and the means of achieving it (which is also the solution to the problem) is the main clause. Thus purpose is a slightly broader concept than problem, while means and solution are the same.

Within the sentence, the Purpose-Means relation is often easier to communicate than Problem-Solution, as we have the handy subordinators *To* (or *In order to*) to indicate purpose, and the equally useful subordinators *by…ing, through…ing* and *by means of* to signal means:

(6) Measure close tolerances by pressing Plasticine into the space and measur*ing* the mould made. (from Jordan, 2000a, p. 104)

Thus, in writing summaries, we can make use of the closeness of these two pairs of relations by expressing Problem-Solution pairs as Purpose-Means sentences or paragraphs in much briefer form within a single sentence.

The concept of purpose, which embraces the objectives, aims, goals, ambitions or desires of those seeking to achieve something, is often a vital piece of information to include in a summary. As the methods, procedures or techniques used to achieve the purpose are often equally important (especially in scientific papers), we also often need to include this sort of information in a summary too. The Purpose-Means relation (e.g., Thompson, 1985; Jordan, 1996; Hwang, 1997) can thus prove to be a powerful guide to what is the most "important" information to include in a summary—as well as providing a method of connecting these two items of information within the summary.

Deletion Heuristics

Most of the discussion so far has dealt with the summarizing technique of compressing information while retaining the essential core meanings and the semantic relations between them. We now need to address the equally important technique of recognizing and deleting less-important types of information in creating the summary. Whether we are creating a summary document from a large amount of information or are summarizing an existing document, we are all familiar with the need to select the most important, useful or relevant items of information for inclusion. While the decisions are often based, sometimes quite subjectively, on our judgment of the needs of our readers and the purpose of the summary, the multi-part and binary relations provide us with a more objective criterion for making the decisions. The related heuristics can prove extremely useful for automatic summarizing, for which total objectivity is the ideal goal.

Working towards providing discourse strategies for the automatic generation of summaries, Rino and Scott (1994, p. 8-10) suggest a series of deletion heuristics based on the binary relations of textual connection. Essentially these mean the deletion of what they regard as the less-important member of the binary pair. Their simple deletions are:

> Delete Particular in the General-Particular relation.
> Delete Elaboration in the Statement-Elaboration relation.
> Delete Assessment in the Statement-Assessment relation.
> Delete Basis from the Assessment-Basis relation.
> Delete Detail from the Preview-Detail relation.
> Delete Example from the Preview-Example relation.
> Delete Background from the Statement-Background relation.

Many of these overlap and are poorly defined, but they do provide a reasonably objective basis for deleting certain types of information. As Rino and Scott recognize that the Problem-Solution relation "constitutes the minimum possible significant pair," they suggest deleting only other parts surrounding that pair. Although they refer to the important Cause-Effect and Purpose-Means relations only peripherally, we should realize that Cause is often less important than Effect, and that Means is often less important than Purpose; these can often be deleted without seriously affecting the value of the summary. Their suggestion of deleting Enabler in the Enabler-Enablement relation is interesting as the Enabler is often a person. More generally, we should recognize that, in technical writing, the human agent is almost always less important

information than the action attributed to that person; this is why the agentless passive is often preferred over the active, which can contain superfluous or even misleading information (Jordan, 1999a, p. 77-78).

Rino and Scott's deletion heuristics should be treated with considerable caution. Sometimes the data themselves (basis) can be more important than the obvious conclusion which can be drawn from them, and occasionally specific information can be more meaningful (certainly often more accurate and reliable) than the generality that can be derived from it. Nevertheless, their approach is a useful one to bear in mind when faced with the task of summarizing information or a document containing binary pairs of relations—as many do. Rino and Scott do not provide useful heuristics for deleting information relating to multi-item structures, perhaps because even semi-objective criteria may not be possible for such complex texts.

Summarizing Descriptive Texts

Descriptive and General Texts

Many texts are essentially "descriptive," meaning that they depend for their cohesion solely on the "re-entry" (Halliday and Hasan, 1976, p. 31; Sidner, 1983, p. 330) of topics into the text (Jordan, 1984a, p. 37); they have few if any Problem-Solution patterns or logical relations of Cause-Effect, Basis-Assessment, Purpose-Means, etc. For such texts, exemplified in "New Products" sections of technical magazines, the strategy for writing summaries relies on the selection of the most important, relevant or interesting bits of information for readers of the text. The only deletion heuristic we can adopt is to delete the "less-important" information, but this advice is too general to be useful. The following section is an attempt to provide some guidelines for the very difficult problem of summarizing descriptive documents or descriptive parts of documents.

While the principles discussed so far in this paper are, at least to some extent, independent of readership, the selection of the "most important" information from general descriptive texts is usually highly dependent on readers. Faced with a list of information about a topic (a narrative about a person, a description of an instrument, details of a news story), the writer needs to decide what readers want (or need) to know most as the basis for selecting the "most important" information That is, the writer needs to answer the question "What is most relevant for my readers?" For descriptive writing, there may be no "gist" in the sense of a sequence of different types of information. But there may be a "kernel"—the major element of information with-

out which the story (or its summary) makes little or no sense. This, of course, must be included in the summary together with other items of information that add to or support the kernel.

Summaries and Relevance Theory

The link has been made recently (Jordan, 2000b) between summaries and relevance theory, a theory which has been hotly debated in cognitive science since its introduction by Sperber and Wilson (1986, 1987). In his chapter on "Description as Summary" and later, Jordan (1984a, p. 8-36) explains relevance of information for technical descriptions first in general terms of the purpose of the description and the needs of readers. He then shows, following Corbett, 1977, p. 41-42, that the "most relevant" material can be generated by asking basic questions about the topic (e.g., "What is it?", "What does it do?", "How does it work?", "What is it used for?", "What does it look like?") as the basis for a summary description. The writer has to select from the information elicited by such questions to determine what is most important for readers.

For many manufactured items, basic engineering functions can be derived from design methodology (e.g., Earle, 1977) such as *measure, move, clean, connect, protect, control, support,* and *sort*) and these then become the basis for "function-based information." This deals with the means, purpose, circumstances, extent, manner and duration of the function—as well as how well the topic performs its function(s). These categories, which coincide with many of the textual relations discussed earlier, then become a useful basis for selecting the most suitable material for the summary based on readers' needs. For example, while engineering readers may need to know the precise level of accuracy of a measuring instrument, students may be more interested in understanding the principle of operation. The systematic approach outlined above provides the writer with a set of types of information, from which a selection can be made for information to be included in the summary.

In light of Sperber and Wilson's definition of relevance as essentially a-contextual (Mey and Talbot, 1988, p. 747) and determinable based on the readers' processing effort, it is ironic that earlier (1982) they had noted that an announcement of **"fire!"** at a theatre presentation was not relevant to the presentation, although it was clearly relevant to the audience! The need for relevance to be related to some defined topic has been discussed by Wilks (1986) and Clark (1987), who both express the view that relevance must be with respect to some defined purpose or goal; Gorayska and Lindsay (1993) and Jordan (1998b) elaborate on their discussion. Wilks (1987) specifically claims that information must also be relevant to a defined person or group, i.e., for an explicit or implicit audience. In spite of the claim by Gorayska and Lindsay (1993, 1995)

that we cannot have degrees of relevance, different levels have been established by Hasan (1985), Martin (1992), Nicolle (1995), and Jordan (1998b). A brief reader-oriented empirical experiment on relevance as the basis for a newspaper summary in Jordan (2000b) notes a general consensus on what material is most relevant, although significant minority views claim other material as more relevant. It appears that the decision as to what information is most relevant in a given situation can be quite objective, but does have some subjective component.

Wider Links with Relevance

Relevance is seen to be a key factor for summarizing by both Rino and Scott (1996) and Robin (1994), as we see in the comparison of methods noted by Rino and Scott (1996, p. 9): "Whereas we identify less *relevant* propositions, omitting them from the primary message source and reorganizing them, Robin identifies *relevant* propositions, adding them to a given proposition in order to build up a text." Although these procedures are opposite in approach, the feature they share is the need to identify *relevant* propositions as the basis for their summary creations. However, what is most relevant for the purpose of a communication is not always obvious. As Winter (1976, preface) warns: "...we cannot rely on intuition or 'doing what comes naturally' when it comes to the precision required in handling complex information which is *relevant* to the purpose of the communication on any scale beyond that which is contained in a mere sentence or two."

Winter's reference to the "purpose of the communication" is noteworthy, as we should not rely solely on readers' needs as the basis for information selection. This broader need for also considering the message the writer wishes to convey is echoed by Rino and Scott (1994, p. 2): "The units are classified according to their *relevance*, and they can be optional or obligatory (i.e., non-essential or essential for the message to be conveyed)." With implicit acceptance of different degrees of relevance, Russell (1979, p. 14) links "relevance" to the concept of "importance" in: "[A]s you analyze the contents weighing the *relevance* of each section and determining how it fits into the general pattern of thought, you will notice that some ideas are of primary importance, and others of secondary importance."

Importance, of course, is a rather vague term, depending on many factors. The statements from Fries (1987, p. 48) and van Dijk (1981, p. 187) noted earlier both include the word "importance," but they make no attempt to define this term in the context of material selection for summaries. Gorayska and Lindsay (1993) suggest the term "interestingness" (see Frick, 1992) as a more appropriate label for what Sperber and Wilson were seeking to describe in their relevance theory. For some texts (e.g., press reports), this may be an appropriate basis for deciding what information to

include in a summary. However, again as noted earlier, for other texts the criterion is better described by terms such as "useful," "stimulating," "informative," "entertaining," "convincing," or "important"—depending on the purpose of the document. That is, the information is important because it is useful, stimulating, informative, etc. Clearly the information selected for inclusion in news report summaries must be interesting, or else they will not be read; but technical reports must primarily be informative, and technical proposals must primarily be convincing. Technical instructions must be useful in helping readers do something, and warnings and cautions are important because they are useful in preventing damage or injury. Perhaps we can best define the overall criterion for summary selection in terms of "what the writer wants readers to know most," and we should be able to apply this to the document as a whole as well as to any summaries of it.

Subjective and Objective Approaches

The approach just outlined is, of course, largely subjective. However, some preliminary empirical evidence (Jordan, 2000a) indicates that there may often be a close consensus regarding which information should be included given the purpose and readership. That is, the intuitive or experienced-based judgments of writers regarding what is the most "important" information in summarizing a document under specified conditions do seem to result in summaries that others would judge to be suitable. For purely descriptive writing, writers may have little more to rely on than their knowledge and experience in this regard.

This subjective approach cannot be used for automatic summary writing, of course, and thus a more objective basis is needed, especially for descriptive texts. The work on lexical connections and structures in texts by Phillips (1983, 1985, 1989) and Hoey (1991) provides a detailed framework for the use of lexical connections as the basis for automatic summary creation. Several other analyses (e.g., Paice, 1990; Francis and Liddy, 1991; Liddy, 1991; and Sparck Jones, 1993) include lexical and syntactical clues, as well as collocation, verb tense distribution and continuity signals, for creating summaries. The automatic summarizing computer tools now available produce interesting and sometimes quite useful summaries, but more work is needed to derive objective heuristics to allow them to create acceptable summaries on all, or even most, occasions. Perhaps their greatest weakness lies in their apparent inability to recognize the kernel of the original material, a matter discussed shortly.

Strategies for Text Summarization

Genre Recognition and Retention

The approach adopted here is that summaries can best be written based on their structural genre, or van Dijk's (1977, p. 137) notion of macrostructure of texts. Once we recognize the overall pattern of the original information, we can duplicate that structure, in miniature form, in a summary that contains the high-priority elements of information. This means that a summary of a Problem-Solution text is still a Problem-Solution text, a summary of a Cause-Effect text is still a Cause-Effect text, a summary of a description is still a description, etc. For reports that provide details, effects, causes, conclusions and recommendations, the summaries will also reflect that sequence and information type although there may be greater emphasis on the conclusions and recommendations. Similarly, summaries for experimental reports that deal with problem statement, method, results, analysis and conclusions will also contain those categories of information, although again with emphasis on the aim, analysis and conclusions.

However, with the exception of some descriptions, we cannot expect the lower structural levels of a document to exhibit the same structure as its macrostructure. For example, a text might have a descriptive macrostructure, but might contain some Problem-Solution or binary logical connections at the section, paragraph or sentence level. Or a text having a Problem-Solution macrostructure might have components that follow a descriptive pattern (usually in the solution) or a logical sequence such as Cause-Effect (usually in the problem). For large texts, there may be several levels of structure related to, but lower than, the levels of their outlines or Table of Contents; and each of these microstructures may have different structures. We thus need to create the summary first based on the macrostructure and then, within the paragraph or section for each major element, based on the microstructures found there.

Noting Russell's (1979, p. 14) comment about "unpacking" vital information for inclusion in the summary, we should be prepared to delete categories that may not be useful for the summary. Occasionally this can be done at the macrostructure level (e.g., the actual data measured in an experimental report), but deletion heuristics are not usually appropriate at the macrostructure level of the text. They are much more useful in deleting sections from within a category of the macrostructure. When that is done, Rino and Scott claim that "the deletion of complex discourse structures... implies the deletion of the related sub-structures" (1994, p. 6). Yet there may be occasions when sub-categories are useful in a summary although higher-level features may be of less importance. For example, within the solution category of a text, per-

haps subcategories of the solution's description, its history, and types available could be deleted, although some brief items in these categories could perhaps still be retained. We see this in Example 7 in the next section.

Identifying and Preserving the Kernel

In addition to preserving the "gist" of the original document (the major content and structural information of the original), we also need to preserve its "kernel." Although several writers point to the need for the "gist" or "essence" to be included in the summary, they are usually less clear about what they mean by these terms. We distinguish here between the "gist" as the set of items of information of different types that collectively convey the "story"—as opposed to the (usually) single, central "kernel" of the story, which is vital to the usefulness or understanding of the text. While it is a premise of Mann and Thompson's "Rhetorical Structure Theory" that there is a central single rhetorical proposition for all texts, the work on problem-solution texts by Winter, Hoey, and Jordan all point to the need for two or more pieces of information to create a "minimum discourse" (Hoey, 1983, p. 31-61) in that genre at least. These views are not inconsistent, as the former involves the kernel, whereas the latter involves the gist. The need to preserve the central feature of an original in a summary is central to Rino's (1996) work on the *Preservação de Idéia Central na Geração de Textos.*

The effect, assessment or purpose of work being summarized could be the gist, but is more likely to need complementary detail to create the gist. The *What?* of a news account is also unlikely to provide an adequate summary on its own; the death of a man in Example 1 by itself needs details of the cause to make it meaningful as a news story, and for this we should probably regard the immediate cause (the accident) and not the actual result as the kernel of the story. Some summaries should include both the gist and the kernel, although some may just have a gist. However, the kernel alone may still be insufficient to create the minimum discourse to stand alone as an effective summary, as we will see in Example 7. If there is a kernel, it *must* be included in the summary and given some prominence as, by definition, it is the information that makes sense of the document; if there are two or more elements which collectively form the gist, they all need to be included even if the summary is very brief.

The kernel is defined here as the central feature on which hinges the understanding, usefulness or newsworthiness of the document. This is seen in the following technical article, which demonstrates principles of both gist and kernel preservation in the summary.

(7) Liquid 'O' Rings

   The two charge machines at Oldbury nuclear power station, near
Bristol, perform all the necessary handling functions associated with
on-load refuelling. On a routine pressure test before a fuelling opera-
tion one of these machines was found to be losing pressure and it was
therefore taken out of service and tested. Ultra-sonic equipment was
used to establish that the problem was in the lower flange of the main
pressure vessel. This is sealed with two 'O' rings, both of which had
failed.

   A major problem arose because the whole of the vessel is encased in
thick shielding; access to the flange could only be made through a spe-
cial inspection plug. Maintenance engineers decided to investigate the
possibility of injecting some kind of plastic material into the interspace
between the 'O' rings and forcing it to spread out along the channel
and seal it.

   The CEGB engineers contacted Sibex (Constructions) Ltd, a com-
pany which specialises in this type of work, and Sibex suggested Devcon
Flexane, which they had used before, as a suitable material.... On cur-
ing, the Flexane assumed the characteristics of resilient rubber.

   After the repair had been completed, the vessel was pressure tested
and found to be completely sealed. (*Chartered Mechanical Engineer*,
September, 1978, p. 28)

   This is obviously a four-part Situation-Problem-Solution-Evaluation structure,
which the writer has obligingly set out in four corresponding paragraphs. The sum-
mary of this summary document should include all four elements of the macrostruc-
ture, of course, to preserve the gist of the story. But perhaps even more importantly,
we need to recognize and preserve the kernel, i.e., its *raison d'être*. This is the *inacces-
sibility* of the failed 'O' rings caused by the shielding necessary for the nuclear plant.
After all, if there were no shielding and the 'O' rings had been accessible, it would have
been a very simple matter indeed to have replaced them—and there would have been
no need for this article, i.e., it would not have been newsworthy. If the first sentence of
the second paragraph were deleted, engineering readers would not understand the
text as a whole; they would not understand why those involved are going to all that
trouble to mend an 'O' ring instead of simply replacing it. The kernel is vital to this
understanding and thus also to the article's usefulness and newsworthiness.

The summary offered for this article in Jordan (1984b, p. 15) retains the crucial kernel of the inaccessibility of the failed parts while also including the gist about all four parts of the Problem-Solution macrostructure:

(8) During a recent routine test at Oldbury nuclear station, a refuelling machine was found to be losing pressure and was taken out of service. The fault was traced to the failure of two inaccessible 'O' rings, and Devcon Flexane was injected around them to provide a seal. On testing, the vessel was found to be completely sealed.

An informative title should also preserve the kernel as well as the gist consisting of the four major elements of the gist: "Liquid Plastic Seals Inaccessible 'O' Rings in Nuclear Plant." This summary-title provides much more information than the original title.

Recognizing and preserving the kernel requires a real understanding of the text and the reason for its usefulness or newsworthiness to the audience; a superficial understanding of the technicalities involved may be insufficient to allow recognition of the kernel in Example 7 and in many other technical examples. The kernel is even more likely to be lost in summaries created by computer automatic summarizing tools, which cannot provide such vital subjective assessment. Although Rino and Scott (1996) identify "gist preservation" as a vital component of effective autosummarizing of original documents, computer tools cannot yet recognize the kernel of the text as they cannot make intelligent decisions about audience understanding, usefulness or newsworthiness. This could result in a poor summary. As an example, even at an 80% summary level for Example 7, the Windows 98 autosummarizer included everything *but* the first sentence of the second paragraph, i.e., it regarded the "kernel" information (which is central to the understanding and newsworthiness of the article) as the "least relevant" part! This serious weakness of autosummarizers must be overcome if they are to produce sound summaries.

Kernel and Audience

A further complication is that the kernel may be dependent on audience, as noted by Rino and Scott (1994, p. 7): "The same discourse component that can be omitted in one context may be obligatory in another, depending on the addressed readership." The kernel for the article in Example 7 is easily recognized by intended readers (mechanical engineers) as it is the item of information that makes sense of the article for them, i.e., without that information the article would have little or no value. However, engineering chemists or materials engineers might also be interested in the material being used and its chemical and mechanical properties on curing; and perhaps

consulting engineers might have a greater interest in the procedure of using special-
ists to solve the problem. Thus by looking at the information from different perspec-
tives, we might find that different parts of the information have greater importance,
and these items (of "auxiliary kernel" information) should then be included in the
summary along with the kernel and the four major categories of information that
constitute the gist.

Many summaries are written for multiple audiences, and these may therefore
need to include different "auxiliary kernels" for these readers—just as a report might
contain information of interest to many different readers. In an industrial accident
report, for example, a common gist would include brief details of the when, where,
and cause(s) of the accident together with information about who was injured and
the extent of their injuries. Yet some readers might like to know about the cost of the
accident, others might be more interested in getting production started again, while
still others would be more interested in any environmental concerns, or perhaps pub-
lic perception. For them, the kernel of the report would be the information of great-
est interest to them. The executive summary, then, might need to include not only the
central gist and kernel of the information, but also the other types of information
which other readers would regard as the kernel for them.

The kernel of an account may not be obvious. We might assume in Example 1
that the central gist is that a man has been killed, yet we can argue that other types of
information make the story more useful or newsworthy. If, for example, the man
killed had been the son of a former Prime Minister of Canada, that small element of
information about the victim becomes central as the story as a whole will command
much greater attention than if another (less notable) individual had been killed. Or if
someone is killed as a result of deliberate shooting in a high school, the location and
circumstances could become the kernel that drives the interest of readers and ana-
lysts. Or if someone is killed during deliberate euthanasia, or in the accidental bomb-
ing of a hospital during war, these become the central idea or kernel that must be
included and emphasized in the summary. Thus, in news reporting in particular, the
basic statement that "A man has been killed." will probably not be the kernel of a
story; the real news value depends on the person killed, the method, the location, the
circumstances, etc. that make up the whole story. The kernel, once recognized, be-
comes the main element of the original story and also of any summaries.

For descriptive texts, and especially for readers of varied interests and back-
grounds, there may be no one kernel or it may contain several elements of informa-
tion. Often no one item of information is essential to the value or usefulness of the
original document as this value could be regarded more as the sum of several parts.
This paper, for example, contains background information, summary strategies for

multi-structured, binary-structured and descriptive texts, explanation of the importance of the gist and kernel, and comments on both human and computer summarizing. Rather than relying on one or more of these items as the kernel for the paper, we are better served by including in the summary brief accounts of all of these major elements of the discussion (the gist) without identifying any one element (the kernel) as being of crucial importance. Readers will no doubt find some sections more or less valuable to them than other sections, i.e., they will have their own kernels.

Dealing with Microstructures

Example 7 also shows us some of the complexity of microstructures in text. The section on Problem includes a decision (Assessment) based on the test results, and the implied Purpose of discovering the Cause of the Problem together with the Means of doing this. The location of the Cause of the Problem becomes the central Problem and a decision (Assessment) is made to investigate (Assess) a certain method (Means) of overcoming this Problem. We also see elements of descriptive text in all parts of the article presented. From these complex connecting systems, we need to extract the four major parts of the story and the central items of information.

The summarizing strategy suggested here is to deal first with the macrostructure of the text—the highest level of the discourse pattern. Once we have established the overall macrostructure in the summary, we concentrate on each of the elements of that structure. Within these, we find other structures, and again we can summarize each of these in turn. However, below the macrostructure, deletion heuristics may be more useful than summarizing, as we "unpack" the more useful information from its surrounding less-useful material; some information at lower levels might be included as a brief note. Although for short original documents this approach could be extended down into the section or possibly the paragraph level, that is less likely to be possible for longer documents in creating a summary of acceptable length. For these, we may have to use only the macrostructure details together with selected information from one or more of the lower levels of document structure.

As noted earlier, descriptive texts occur as macrostructures, or as microstructures within one or more of the major structures of information of the macrostructure. The procedure for dealing with descriptive texts discussed here applies whether the text itself is descriptive or just parts of it are. For description as a macrostructure, we first need to identify and include in the summary the kernel (if there is one) and all other major groups of information in the original material; then we can look within these groups to see if there is any further useful information for inclusion. Whenever descriptive text occurs in microstructures, we may be able to delete it—especially the more specific elements of General-Particular, General-Example, Part-Whole, Abstract-

Instance or Unspecific-Specific groupings. Remember, though, that some such specific information could form part of the kernel or auxiliary kernel of the whole message and should therefore be included in brief form.

Conclusions

The Strategies

This paper has explained techniques for using the structures of information in the original material for creating summaries. At this level, the approach is largely objective and thus is applicable for any computer summarizing tool that can recognize the different types of information and their related signals. The Problem-Solution macrostructure is taken as an example of how to use high-priority items of information from the elements of multi-item texts as the basis for the different successively-briefer levels of summary, including very brief introductory summaries and even titles.

Binary relations are also discussed as the basis for creating summaries based on these macrostructures of original material. As macrostructures, the logical relations of Cause-Effect, Assessment-Basis and Purpose-Means are shown to be useful in understanding the gist of original material, as are the more particular binary relations of Part-Whole, General-Specific, etc. As microstructures, they form the basis for deletion heuristics, in which one element of the pair is perceived as being less essential than the other—and thus candidates for deletion in creating the summary.

Descriptive texts and information cannot be summarized on the basis of their information structure like multi-item and binary texts. For these, the selection of material for inclusion in the summary is based on answers to series of questions relating to identifiable categories of information specific for different types of subject. The question of relevance, audience need and document purpose become much more important as criteria for selecting the most useful material, and this involves more-subjective judgments than we find for the more-structured texts.

Rules and Kernel

These strategies are all useful techniques, but Ratteray (1985, p. 469) warns of the dangers of applying such sets of rules of summarizing too rigidly without considering the overall concept of the document being summarized. He also suggests that summarizers develop an understanding of summary types to avoid confusing the criteria for creating a summary. These cautions are worth noting in connection with the methods advocated here. The method of using the contents and structure of the original material as the basis for a summary is well founded, but an overly-formulaic

approach could well result in inappropriate summaries. Use of the concept of the "kernel" as well the "gist" of original material should help summarizers to steer clear of summaries that may provide a readable and quite representative summary structurally while missing the whole point (the kernel) of the original work.

For autosummarizing, however, we must use sets of rules, for these are all computers can understand. The signalling of *A major problem* in Example 7 should be clear enough to attract the attention of most human summarizers and we should, in our writing, make the kernel abundantly clear to highlight such essential information. Even so, computers may find such signals difficult to recognize as indicators of kernel. Work on this aspect of computer summarizing is needed to complement the useful work already done.

Autosummarizing tools are also extremely weak in creating short summaries for lengthy documents. The 10-sentence computer summary for this paper, for example, is a garbled mess of headings in non-sentence form. Although such a "summary" may be useful in informing readers of some of the contents of the paper, it is unacceptable as a preceding abstract, which is traditionally in sentence form. While such tools may prove useful in creating initial summaries at higher percentages of the original (see Jordan, 2000b), significant refinements are needed before they can produce useful brief summaries of large documents.

REFERENCES

Baker, E. L., N. K. Atwood & T. M. Duffy. (1988). Cognitive approaches to assessing the readability of texts. In A. Davidson and G. M. Green (Eds.), *Linguistic complexity and text comprehension: Readability issues reconsidered* (pp. 55-83). Hillsdale, NJ: Lawrence Erlbaum.

Basham, C. S. (1986). *Summary-writing: A study in textual and contextual constraints.* Doctoral dissertation, Michigan University, 8621250.

Beaudet, C. (1994). Pour une typologie des résumés fonctionnels, *Technostyle,* 11 (3/4), 49-58.

Beardsley, M. C. (1950). *Practical logic.* New York: Prentice Hall.

Beekman, J., & J. C. Callow. (1974). *Translating the word of God.* Grand Rapids, MI: Zondervan.

Blicq, R. S. (1983). *Technically-write.* Scarborough, ON: Prentice Hall.

Boret, M. & J. Peyrot. (1971). *Le résumé de texte.* Paris : Chotard et Associés.

Bower, G. H., & R. K. Cirilo. (1985). Cognitive psychology and text processing. In
    van Dijk (Ed.), *Semantic discourse analysis* (pp. 71-105). London: Academic
    Press, Harcourt Brace Jovanovich.

Brown, A. L. & J. D. Day. (1983). Macrorules for summarizing texts: The
    development of expertise. *Journal of Verbal Learning and Verbal Behavior, 22,*
    1-14.

Burke, K. (1969). *A grammar of motives.* Berkeley and Los Angeles: University of
    California Press (originally published 1945).

Clark, H. H. (1987). Relevance to what? *Journal of Behavioral and Brain Sciences,*
    10(4), 714-715.

Collison, R. L. (1971). *Abstracts and abstracting services.* Santa Barbara, CA:
    American Bibliographic Center.

Corbett, E. P. J. (1977). *The little rhetoric and handbook.* New York: John Wiley and
    Sons.

Cremmins, E. T. (1982). *The art of abstracting.* Philadelphia: ISI Press.

Earle, J. H. (1977). *Engineering design graphics.* Reading, MA: Addison Wesley.

Francis, H., & E. D. Liddy. (1991). Structured representation of theoretical abstracts:
    Implications for user interface design. In M. Dillon (Ed.), *Interfaces for
    information retrieval and online systems: The state of the art* (pp. 107-114).
    Westport, CT: Greenwood Press.

Frick, R. W. (1992). Interestingness. *British Journal of Psychology,* 83(1), 113-128.

Fries, U. (1987). Summaries in newspapers: A textlinguistic investigation. In U. Fries
    (Ed.), *The structure of texts* (pp. 47-63). Swiss Papers in English Language and
    Literature, Tübingen: Gunter Narr Verlag.

Gorayska, B., & R. Lindsay. (1993). *The roots of relevance.* Journal of Pragmatics, 19,
    301-323.

——— (1995). Not really a reply – More like an echo: Reply to Steve Nicholle, *Journal
    of Pragmatics,* 23, 683-686.

Halliday, M. A. K. (1978). *Language as social semiotic: The social interpretation of
    language and meaning.* London: Edward Arnold.

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

Hasan, R. (1985). The structure of a text. In M. A. K. Halliday & R. Hasan (Eds.), *Language context and text: Aspects of language in a social-semiotic perspective* (pp. 70-96). Geelong, Victoria, Australia: Deakin University Press [Oxford University Press, 1989].

Hidi, S. (1983). *Summarization of complex texts.* Toronto: Ontario Institute of Studies in Education.

Hobbs, J. R. (1985). *On the coherence and structure of discourse.* Technical Report CSLI-85-37, Stanford University, Center for the Study of Language and Information.

Hoey, M. P. (1979). *Signalling in discourse.* Discourse Analysis Monographs No. 6, Birmingham, England: English Language Research, University of Birmingham.

—— (1983). *On the surface of discourse.* London: George Allen and Unwin.

—— (1991). *Patterns of lexis in text.* Oxford: Oxford University Press.

Hovy, E. H. (1990). Unresolved issues in paragraph planning. In R. Dale, C. S. Mellish & M. Zock (Eds.), *Current research in natural language generation* (pp. 17-45). London: Academic Press.

Hutchins, W. J. (1977a). On the surface of scientific discourse. *UEA Papers in Linguistics,* 5 (3), Norwich, Norfolk: University of East Anglia, 18-39.

—— (1977b). On the problem of "aboutness" in document analysis. *Journal of Informatics,* 1(1), 17-33.

Hwang, S. J. J. (1997). Purpose clauses in English and Korean. In A. K. Melby (Ed.), *The 24th LACUS forum,* 496-508, Chapel Hill, NC: LACUS.

Johns, A. M. (1985). Summary protocols of 'underprepared' and 'adept' university students. *Language Learning,* 35(45), 495-512.

Jordan, M. P. (1980). Short texts to explain problem-solution structures—and vice versa. *Instructional Science,* 9, 221-252.

—— (1981). Structure, meaning and information structures of some very short texts. In W. Gutwinski & G. Joly (Eds.), *11th LACUS forum* (pp. 410-417). Columbia, SC: Hornbeam Press.

—— (1984a). *Fundamentals of technical description.* Malabar, FA: Krieger.

—— (1984b). *Rhetoric of everyday English texts*. London: George Allen and Unwin.

—— (1988). How can problem-solution structures help writers plan and write technical documents? In L. Beene & P. White (Eds.), *Solving problems in technical writing* (pp. 3-26). Oxford: Oxford University Press.

—— (1991a). The linguistic genre of abstracts. In A. M. Della Volpe (Ed.), *The 17th LACUS forum* (pp. 507-527). Lake Bluff, IL: LACUS.

—— (1991b). Openings in informal scientific texts. *Technostyle*, 9(2), 1-19.

—— (1991c). Openings in quite formal scientific texts. *Technostyle*, 9(3), 18-37.

—— (1993). Openings in formal scientific and engineering texts. *Technostyle*, 11(1), 1-28.

—— (1996). To-infinitive groups as clauses and post-modifiers. In M. -J. Powell (Ed.), *The 21st LACUS forum* (pp. 258-268). Lake Bluffs, IL: LACUS.

—— (1997). Subtle signalling of logical relations in news reporting. In A. K. Melby (Ed.), *The 23rd LACUS forum* (pp. 325-340). Chapel Hill, NC: LACUS.

—— (1998a). Pragmatic, stylistic and grammatical limitations on choice: A study of cause-effect relations in English. In A. Sánchez-Macarro & R. Carter (Eds.) *Linguistic choice across genres: Variation in spoken and written English* (pp. 65-86). Amsterdam: John Benjamins.

—— (1998b). The power of negation: Text, context and relevance. *Journal of Pragmatics*, 29, 705-752.

—— (1998c). Enabler-enablement relations in English. In S. Embleton (Ed.), *The 24th LACUS forum* (pp. 290-303). Chapel Hill, NC: LACUS.

—— (1999a). "Unattached" clauses in technical writing. *Journal of Technical Writing and Communication*, 29 (1), 65-93.

—— (1999b). Towards a rhetoric of informal scientific writing: Plausible argument and complex reasoning. *Technostyle*, 15(1), 1-19.

—— (2000a). *The language of technical communication*. (2 ed.) Kingston, ON: Quarry Press.

—— (2000b). A pragmatic/structural approach to relevance. *Technostyle*, 16(2), 47-67.

—— (2001). Some discourse patterns and signalling of the assessment-basis relation. In M. Scott & G. Thompson (Eds.), *Patterns in text* (pp. 159-192). Amsterdam: John Benjamins,

Katz, M. J. (1985). *Elements of the scientific paper.* New Haven, CT: Yale University Press.

Kintsch, W. (1985). Text processing: A psychological model. In T. van Dijk (Ed.), *Semantic discourse analysis* (pp. 231-257). London: Academic Press, Harcourt Brace Jovanovich.

Kintsch, W., & T. A. van Dijk. (1978). Toward a model of discourse comprehension and production. *Psychological Review, 85,* 363-394.

Labov, W. (1972). *Language of the inner city: Studies on the black English vernacular.* Philadelphia: University of Pennsylvania Press.

Labov, W. & J. Waletsky. (1967). Narrative analysis: Oral versions of personal experience. In *Essays on the verbal and visual arts* (pp. 12-44). Seattle, WA: University of Washington Press.

Laurent, J. P. (1985). L'apprentissage de l'acte de résumer. *Pratiques, 48,* 71-89.

Liddy, E. D. (1991). The discourse-level structure of empirical abstracts: An exploratory study. *Information Processing and Management, 27*(1), 55-81.

Longacre, R. E. (1972). *Hierarchy and universality of discourse constituents in New Guinea languages.* Georgetown: Georgetown University Press.

Maizell, R. E., J. F. Smith & T. E. R. Singer. (1971). *Abstracting scientific and technical literature.* New York: Wiley-Interscience, John Wiley and Sons.

Mann, W. C., & S. A. Thompson. (1987). *Rhetorical structure theory: A framework for the analysis of texts.* Technical Report ISI/RS-87-185, Marina del Rey, CA: University of Southern California. Also in *Discourse Processes, 9*(1), 57-90.

—— (1989). *Rhetorical structure theory: A theory of text organization.* Technical Report ISI/RS-87-190, Marina del Rey, CA: University of Southern California.

Manning, A. D. (1990). Abstracts in relation to larger and smaller discourse structures. *Journal of Technical Writing and Communication, 20*(4), 369-390.

Martin, J. (1992). *English text: System and structure.* Amsterdam: John Benjamins.

Mey, J. L. & M. Talbot. (1988). Computation and the soul. *Journal of Pragmatics*, 12,
    743-789.

Moreau, J. A. (1981*). La contraction et la synthèse de textes*. Paris : Éditions Fernand
    Nathan.

Nicolle, S. (1995). In defence of relevance theory: A belated reply to Gorayska and
    Lindsay, and Jucker. *Journal of Pragmatics*, 23, 677-681.

Paice, C. D. (1990). Constructing literature abstracts by computer: Techniques and
    prospects. *Information Processing and Management*, 26(1), 171-186.

Pearson, P. D., & K. Camperell. (1981). Comprehension of text structures. In J. T.
    Guthrie (Ed.), *Comprehension and teaching: Research reviews* (pp. 27-55).
    Newark, DE: International Reading Association.

Pechenik, J. A. (1993). *A short guide to writing about biology*. New York: Harper
    Collins.

Peirce, C. S. (1978). Ideas, stray or stolen, about scientific writing, No. 1. *Philosophy
    and Rhetoric*, 11(3), 147-155.

Phillips, M. (1983). *Lexical macrostructure in science text*. Doctoral dissertation,
    University of Birmingham, England.

—— (1985). *Aspects of text structure: An investigation of the lexical organisation of
    texts*. Amsterdam: North-Holland.

—— (1989). *The lexical structure of text*. Discourse Monographs, No. 12.
    Birmingham England: English Language Research Centre, University of
    Birmingham.

Ratteray, O. M. T. (1985). Expanding roles for summarized information. *Written
    Communication*, 2(4), 457-472.

Rino, L. H. M. (1996). *Modelagem de discurso para o tratamento da concisão e
    preservação de idéia central na geração de textos*. Doctoral dissertation,
    Universidade de São Paulo, Brazil.

Rino, L. H. M., & D. Scott. (1994). *Content selection in summary generation*, Report
    ITRI-94-9, Technical Report Series, Brighton, Sussex, England: Information
    Technology Research Institute.

—— (1996). *A discourse model for gist preservation*, XIII Simpósia Brasileiro de
    Inteligencia Artificial.

Robin, J. (1994). *Revision-based generation of natural language summaries providing historical backgrounds: Corpus-based analysis, design, implementation and evaluation.* Doctoral dissertation, Columbia University.

Russell, P. (1989). *How to write a précis.* Ottawa: University of Ottawa Press.

—— (1992). *The integration of theory and practice in the development of summary-writing strategies.* Doctoral dissertation, Université de Montréal.

—— (1994). Investigating summary typology: Considerations for classification. *Technostyle,* 11(3/4), 37-47.

Sherrard, C. (1989). Teaching students to summarize: Applying textlinguistics. *System,* 17(1), 1-11.

Sidner, C. L. (1983). Focusing in the comprehension of definite anaphora. In M. Brady & R. C. Berwick (Eds.), *Computational models of discourse* (pp. 267-330). Cambridge, MA: The MIT Press.

Sparck Jones, K. (1993). *Discourse modeling for automatic summarising.* Technical Report No. 290, Cambridge: Cambridge University Press.

Sperber, D., & D. Wilson. (1982). Mutual knowledge and relevance in theories of comprehension. In N. V. Smith (Ed.), *Mutual knowledge* (pp. 61-85). London: Academic Press.

—— (1986). *Relevance: communication and cognition.* Oxford: Basil Blackwell.

—— (1987). Précis of relevance: Communication and cognition. *Behavioural and Brain Sciences,* 10: 695-754.

Sweet, H. (1891). *A new English grammar.* Oxford: Oxford University Press.

Thompson, S. A. (1985). Grammar and written discourse: Initial and final purpose clauses in English. *Text,* 5, 55-84.

Valentine, E. (1990). *Précis-writing,* Montréal : Sodilis.

van Dijk, T. A. (1977). *Text and context: Explorations in the semantics and pragmatics of discourse.* London: Longman.

—— (1981). *Studies in the pragmatics of discourse.* The Hague: Mouton.

—— (1985). *Semantic discourse analysis.* London: Academic Press, Harcourt Brace Jovanovich.

van Dijk, T. A., & W. Kintsch. (1983). *Strategies of discourse comprehension.* New York: Academic Press.

Werlich, E. (1976). *A text grammar of English.* Uni-Taschenbücher 597, Heidelberg: Quelle and Meyer.

Wilks, Y. (1986). Relevance and beliefs. In T. Myers, K. Brown & B. McGonigle (Eds.), *Reasoning and discourse processes* (pp. 265-289). New York: Academic Press.

—— (1987). Relevance must be to someone. *Journal of Behavioral Science,* 10: 735-736.

Winograd, P. N. (1984). Strategic difficulties in summarizing texts. *Reading Research Quarterly,* 19, 404-424.

Winter, E. O. O. (1976). *Fundamentals of information structure.* Hatfield, Herts: School of Humanities, The Hatfield Polytechnic (now Hertfordshire University).

—— (1992). The notion of unspecific versus specific as one way of analysing the information of a fund-raising letter. In W. C. Mann & S. A. Thompson (Eds.), *Discourse description: Diverse linguistic analyses of a fund-raising text* (pp. 131-170). Amsterdam: John Benjamins.

Young, R. E, & A. L. Becker. (1966). The role of lexical and grammatical cues in paragraph recognition. In *Studies in language and language behavior,* Progress Report No. 2, Ann Arbor, MI: University of Michigan, Center for Research on Language.