Generation by selection and repair as a method for adapting text for the individual reader*

Chrysanne DiMarco
Department of Computer Science
University of Waterloo
Waterloo, Ontario N2L 3G1, Canada

Graeme Hirst
Department of Computer Science
University of Toronto
Toronto, Ontario M5S 3G4, Canada

Eduard Hovy
USC / Information Sciences Institute
4676 Admiralty Way
Marina del Rey, CA 90292-6695, U.S.A.

Introduction

A recent and growing development in Web applications has been the advent of various tools that claim to "customize" access to information on the Web by allowing users to specify the kinds of information they want to receive without having to search for it or sift through masses of irrelevant material. But this kind of customization is really just a crude filtering of raw Web material in which the user simply selects the "channels" of information she wishes to receive; this selection of information sources is hardly more "customization" than someone deciding to tune their television to a certain station. True customization, or tailoring, of information would be done *for* the user by a system that had access to an actual *model of the user*, a profile of the user's interests and characteristics. And such tailoring would involve much more than just selecting streams of basic content: the content of the text, whether for an on-line Web page or a paper document, would be carefully selected, structured, and presented in the manner best calculated to appeal to a particular individual. Adaptive-hypertext presentation comes closest to achieving this kind of document tailoring, but the current techniques used for adapting the content of a document to a particular user generally only involve some form of selectively showing (or hiding) portions of text or choosing whole variants of larger parts of the document.[1]

If the Web document designer wishes to write and present material in a way that will communicate well with the user, then just displaying the most relevant chunks of information will not be sufficient. For effective communication, both the form and content of the language used in a document should be tailored in rhetorically significant ways to best suit a user's particular personal characteristics and preferences. Ideally, we would have Web-based natural language generation systems that could produce fully customized and customizable documents on demand by individual users, according to a formal user model. As a first step in this direction, we have been investigating applications of our earlier work on pragmatics in natural language processing to building systems for the automated generation of Web documents tailored to the individual reader.

The HealthDoc approach to automated generation of tailored natural language documents

The HealthDoc project

Our long-term research goal is to develop theories of text composition—in particular, computational models of rhetorical structure, lexical semantics, and fine-grained meaning—which are particular requirements for incorporating style and rhetoric in natural language systems. Our current focus of research is our HealthDoc project (DiMarco, Hirst, Wanner, and Wilkinson 1995), which is developing natural language generation systems for producing health-information and patient-education documents that are customized to the personal and medical characteristics of the individual patient. The HealthDoc approach is applicable, we believe, to many kinds of situation in which the ability to target tailored documents to the characteristics of specific users would be desirable. This kind of customization would involve much more than just producing each document in half a dozen

*This paper can be accessed from: http://logos.uwaterloo.ca/∼healthdo/Publications/hypertext97.html.

[1] Brusilovsky (1996) describes such techniques for adaptive hypertext presentation as *conditional text*, selecting a chunk of text based on an associated condition, *stretchtext*, 'collapsing' or uncollapsing textual extensions at certain points in the document, and *frame-based presentation*, using a frame representation to store related textual variants and links to other relevant information.

different versions for different audiences. Rather, the number of different combinations of factors might easily be in the tens or hundreds of thousands, making it impossible to produce, in advance of need, the large number of different versions of each document that would be entailed by individual tailoring of information. This is exactly the situation that we face in developing Web-based natural language generation systems that could produce tailored documents, whether ordinary text or hypertext, for the individual Web user.

Recently, we have been experimenting with the application of HealthDoc techniques to develop a related system, WebbeDoc (DiMarco and Foster 1997), that can customize Web documents on demand, according to a profile of an individual user. In the sections below, we describe the first WebbeDoc prototype, then outline the kinds of long-term research issues which will need to be addressed to develop a full working system. To start, we present an overview of the concepts and techniques that HealthDoc uses and that are being adapted for WebbeDoc.

The master document and generation by selection and repair

The key idea in the HealthDoc approach to producing tailored documents is that we start from an existing *master document* that is then customized for a particular audience. A master document is an encapsulation of all the variations on a given topic that might be needed for any potential reader; it is represented in an abstract, albeit language-dependent, text specification language that expresses not only the content of the document but also information that will assist any subsequent process of revision; this language will be described below. Selections from this document are made for both content and form, but are automatically post-edited—"*repaired*"— for form, style, and coherence. A master document, when its text has been particularized for an individual reader, can be disseminated in a variety of forms; for example, paper, Web page, Web hypertext document.[2]

How repairs are made

The core of HealthDoc's tailoring facility is its *sentence planner*, which is presently under development for the main project. Because the bits and pieces of text selected from a master document might not necessarily be coherent or cohesive, the sentence planner performs complex linguistic repairs to restore coherence and cohesion to the "broken" selected document.

The sentence planner takes as input a set of sentence plans, written in Text Specification Language (described below), and performs the necessary repairs, with each type of repair performed by an independent repair module. The sentence planner is based on a blackboard architecture in which individual repair modules communicate and resolve their conflicts with one another. The architecture is described in greater detail by Hovy and Wanner (1996) and Wanner and Hovy (1996). Four repair modules are being built in the first phase of the main HealthDoc project: for discourse structuring and rhetorical relations, aggregation to remove redundancies, reference restoration using pronouns, and constituent re-ordering.

The first phase: Text generation by selection only in the WebbeDoc system

We have developed a first application of the ideas and methods of HealthDoc in a prototype Web-based system, called WebbeDoc (DiMarco and Foster 1997), that tailors a Web document describing the HealthDoc project to the individual reader.[3] WebbeDoc first displays only the most basic information about the project, using a bland style of presentation, and then allows the user to set various personal parameters and stylistic preferences. This causes WebbeDoc to "rewrite" its text and presentation style in accordance with the selected reader profile. Users can specify their role (e.g., computational linguist, layperson), age, and reading level, as well as stylistic preferences about the formality or "coolness" of the document to be generated.[4] Examples of three different tailored versions of the opening section of the WebbeDoc page are shown in figures 1 to 4.

A document can be customized on all levels of linguistic structure: paragraph, sentence, and lexical choice, with each type of structure chosen for the appropriate pragmatic effect. WebbeDoc is doing more than blindly concatenating blocks of information content; it is selecting the most relevant pieces of text, with respect to both semantic content and pragmatic effect, so that they fit together in a coherent and cohesive manner. It is the existence of explicit rhetorical and other linguistic relations between the individual pieces of text that gives WebbeDoc the ability to produce coherent and polished tailored documents from the master document.

The document's structural representation contains not only linguistic information, but formatting specifications for each type of reader. So, in addition to

[2]We are also experimenting with the inclusion of non-textual variations in a master document that, like ordinary text, may be selected for a particular reader. Some examples from our customizable project home page showing tailoring of illustrations, as well as text, are given in figures 1 to 4.

[3]A demonstration of WebbeDoc can be found at: http://logos.uwaterloo.ca/~healthdo/About/Demos/webbedoc.html.

[4]Currently, WebbeDoc uses a total of five reader parameters: role (computational linguist, physician, funder, layperson); degree of technical detail (high, low); degree of formality (formal, informal); age (child, adult, senior), and degree of "coolness" (bland, cool). This gives a possible 96 ($4 \times 2 \times 2 \times 3 \times 2$) distinct combinations and different texts.

Figure 1: The HealthDoc home page, tailored by WebbeDoc for a non-technical layperson

Figure 2: The HealthDoc home page, tailored by WebbeDoc for a highly technical layperson

Figure 3: The HealthDoc home page, tailored by WebbeDoc for a project funder

Figure 4: The HealthDoc home page, tailored by WebbeDoc for a computational linguist

textual customization, WebbeDoc can tailor the document's style and form of presentation; it can select, according to the user profile, the most appropriate artwork, font, colour, and general layout.

A WebbeDoc master document can also incorporate hypertext variations, which can be embedded within larger textual variations, or linked to specific lexical variations. For example, WebbeDoc can select from among a set of near-synonyms according to the user profile, with each near-synonym linked to a stylistically and semantically appropriate hypertext variation.

WebbeDoc represents the first phase in the implementation of the ideas and mechanisms of HealthDoc. The project Web page that it customizes is itself a master document, but in this initial implementation, we have implemented a form of "generation by selection only": the structure of the master document is tightly constrained so that, after selection, no repairs will be needed to produce a coherent and stylistically adequate text.

The key to WebbeDoc's ability to produce tailored documents by selection from a single master document is the manner of representation of the master document: a WebbeDoc master document has a well-defined structure of ordering relations, rhetorical relations, and other linguistic information, such as coreference links. In the first prototype, the master document was built manually according to our model of a master document, with additional structural constraints imposed so that piecewise selection and recombination would not create any infelicities such as abrupt changes of topic, unnecessary duplications of noun phrases, or unresolvable pronouns.

But to compose a master document of this style and internal complexity required the efforts of computational linguists, rhetoricians, and Web document designers; obviously this is not realistic for the average Web user! In a realistic and usable implementation, WebbeDoc would need an authoring tool and a sentence planner that could work in real-time to repair and polish the selected text—we can't expect the average Web document author to pre-compile all the possible combinations in advance. Therefore, to develop such a system, a number of research issues must be addressed, including representation of the master document; authoring and knowledge-based document management; and sentence planning for automated post-editing.

The next step: Adapting Web pages for the individual user by selecting and repairing text

Representing a master document

Text Specification Language, or TSL, is the language used to represent master documents in the parent HealthDoc system. We anticipate that WebbeDoc master documents will have a hybrid representation: part TSL (for the portions that will be subject to syntactic or stylistic repair), and part "frozen" English text (for the portions that need never be revised). We have defined TSL to incorporate structures represented in Sentence Plan Language (SPL), the specification language for the Penman text generation system (Penman Natural Language Group 1989), whose KPML derivation (Bateman 1995) is used in HealthDoc.[5] An SPL expression is an abstract specification of a sentence, which Penman can convert to the corresponding surface form. This permits expression of the content of the document. The basic SPL structures are annotated

[5]TSL can actually incorporate multiple representations of a sentence; for example, the WebbeDoc system currently uses TSL with both English and HTML representations.

with information for selection and repair to produce the corresponding TSL representation.

The selection information makes reference to a general user profile that describes the possible characteristics of the potential readers: a list of all selection features is stored with their set of possible values. For example, the personal characteristics of WebbeDoc readers might be decribed by a set of selection features and possible values as follows:

```
:reader-role (layperson physician computer-expert)
:reader-age (child adult senior)
```

Other kinds of selection features, such as reading level and preferred style of presentation, will, for the moment, be represented in a similar manner:

```
:technical-level (low-technical high-technical)
:formality (informal informal)
```

A selection condition is a boolean expression composed of particular values of selection features; for example:

```
:condition '(AND (OR layperson physician)
                 low-technical))
```

Such selection conditions can be included as annotations at any level in the TSL so that the system can make selections at any level of linguistic granularity.[6]

But this information isn't enough. We also require the internal discourse structure to be represented explicitly, to guide repairs to the structure of the text. Therefore, TSL contains several kinds of additional annotations, including *topic ordering information*, *coreference links*, and *rhetorical relations* between sentences. As stylistic and pragmatic customization becomes more complex, additional representations will probably be needed. In addition to these kinds of annotations, WebbeDoc's TSL will contain information on formatting and document presentation that would be marked up for inclusion according to specific user preferences.[7]

The model of a master document A master document is constructed according to a formal model; the model that we describe here is the most general, intended for the overall HealthDoc system, which does selection and repair of a master document. (The current version of WebbeDoc, which does generation by selection only, with no repairs involved, uses a more constrained model of a master document.)

We define the general model of a master document (MD) as follows:

- An MD has a coherent high-level communicative goal, such as to inform, to command, to persuade, to impress. For example, the purpose of the current WebbeDoc MD is to inform (and impress) the reader about the goals and technical achievements of the HealthDoc project.
- An MD has a coherent topic structure, with a division into topics, sub-topics, and so on. The smallest topic unit of an MD at the moment is a sub-sub-topic; however, we believe the form of the "smallest topic unit" will vary with the particular document. For example, a master document giving someone information on the treatment of their diabetes[8] might start with a definition of the two different types of diabetes, followed by the identification of the reader's particular type of diabetes has and then a description of the medical characteristics of the two kinds of diabetes.
- Each sub-topic corresponds to a section of the document that satisfies a more specific communicative goal, such as to justify or elaborate upon. In the diabetes example, one sub-topic elaborates on the two types of diabetes, first identifying the reader's particular form of diabetes then describing the characteristics of the two different forms. Essentially, a sub-topic is a semantically coherent piece of the document.
- Each sub-topic is a collection of *variation sets* that are connected by ordering relations, rhetorical relations, coreference links, and formatting relations. A variation set is a set of textual variations such that each variation fulfills the same communicative goal, but has a semantic content and pragmatic form tailored to a particular audience. Each variation in a variation set is characterized by a logical condition and a semantically coherent piece of text. The logical condition uses terms that range over sets of mutually exclusive features. We interpret "mutual exclusion" to mean that the conditions assigned to the variations in a variation set define a clean partition of the set, so that exactly one of the variations must be chosen.

  In the diabetes example, the sub-topic on the two types of diabetes might contain a sequence of two variation sets, the first identifying which condition the reader has and the second explaining the medical details of the different conditions.
- *Ordering relations* may exist between the variation sets that make up a sub-topic. These relations indicate the preferred order of the sequence of variations that have been selected to form the working document, and thereby specify the ordering of

[6]The tradeoffs between amount of variation, grain size of variation, and effort of document authoring and repair are a matter for empirical investigation, and will eventually constitute one of HealthDoc's theoretical contributions.

[7]Indeed, we anticipate that there will be a distinct "repair" module for document formatting in the sentence planner used with WebbeDoc.

[8]An example master document that gives basic information on the treatment of diabetes can be found at: http://logos.uwaterloo.ca/~healthdo/About/Demos/diabetes.html.

sub-topics prior to the invocation of the sentence planner.

Preferred order can vary by reader. In our example, we've assumed that the order of topics should be to first identify the reader's particular type of diabetes, then elaborate on the medical characteristics of the relevant type of diabetes. However, some readers might prefer to be presented first with the overall description of the two different conditions before focussing on their specific problem.

- *Rhetorical relations* may exist between the variation sets that make up a sub-topic. The rhetorical relations that we are currently using are taken from Rhetorical Structure Theory (RST) (Mann and Thompson 1988). In the current version of WebbeDoc, the *same* rhetorical relation must exist between any two members of adjacent variation sets. In the example we have been using, any choice from the variation set describing the medical details of each form of diabetes would have the same rhetorical relation, *elaboration*, to any choice from the first variation set, which identifies the particular form that the reader has.
- *Coreference links* may be defined between any two variation sets. A sequence of coference links in the diabetes document could include: *diabetes*, *insulin-dependent diabetes*, *your condition*.
- *Formatting information* may be defined at each topic and sub-topic level. Formatting information may also be defined between and within variation sets, including illustrations, choice of colour, design of layout, and so on.

Authoring a master document

WebbeDoc master documents may be based on the natural-language text of pre-existing material, or they may be created from scratch (or some combination of the two). Either alternative requires the involvement of a human.

The author of a WebbeDoc master document would normally be a professional technical writer or Web-document designer, who will need to understand the nature of customized and customizable texts, but who should not be assumed to have any special knowledge or understanding of TSL or the innards of WebbeDoc. The authoring tool, therefore, should be no more difficult for the author to use than, say, the more-sophisticated features of a typical word processor. The text is therefore written in English, and will be translated to TSL by the authoring tool. (The English source text is retained in the TSL for use in subsequent authoring sessions—for example, if the document is updated or amended.)

It is the writer's job to decide upon the basic elements of the text, the formatting, ordering, rhetorical, and coreferential links between them, and the conditions under which each element should be included in the output. The elements of the text are then typed into the authoring tool in English, and are marked up by the writer with conditions for inclusion, links for cohesion and coreference, and annotations for ordering and formatting of the document layout.

The tool then translates the text into TSL, including the conversion of the English text into SPL. The latter process is essentially one of semi-automated parsing, so that whenever an ambiguity cannot be resolved, the writer is queried in an easy-to-understand form. The design and development of the authoring tool and its user interface is part of the current phase of the overall HealthDoc project (fall 1996 to spring 1997). The user interface is being developed by Parsons (1997), while Banks (1997) is implementing the English-to-SPL conversion (for more details on the underlying model of conversion, see DiMarco and Banks (1997)).

Functions of sentence planning and automated post-editing

In general, selecting material from pre-existing text and then editing it to recover coherence and cohesion can involve a wide range of problems in various aspects of sentence planning. For example, both syntactic and semantic aggregation may be needed, as well as chunking of whole and partial propositions. Pronouns and other forms of reference need to be chosen. And, of course, aggregation and sentence restructuring will affect the rhetorical relations between the elements of the text.

Our current work is focusing on the development of two key modules of the sentence planner: for discourse structuring and for aggregation.

It is unlikely that every ordering of the blocks of text that are organized into a master document will produced a coherent sequence of selected pieces of text. To ensure that any resulting document makes sense, the discourse structuring module uses the rhetorical relations that hold among the textual units to produce a sequence that is most likely to be coherent. Its rules are derived from a set of Rhetorical Structure Theory relations (Mann and Thompson 1988) whose Nucleus and Satellite ordering requirements are implemented as constraints; the module applies a constraint satisfaction algorithm to find all satisfactory ordering(s) of the input expressions. Of these, one is selected at random. See Marcu (1997) for details. In later work, an additional module will be built to determine the linguistic phrasing of the discourse relation.

The aggregation module eliminates redundancy in TSL expressions by grouping together entities that are arguments of the same rhetorical relation, verbal process, etc. Each aggregation rule recognizes an exact match of some portions of two input TSL expressions and returns a single, fused, expression. The actions of the aggregation module will generally affect the resulting syntactic structure.

A critical problem is the distribution of repair tasks among the planning modules, as there are often strong interactions. The responsibilities of each module and the overlaps between them are an area of on-going research for our sentence-planning group.

Related work

A number of other projects have also used a combination of natural language generation techniques and hypertext capabilities to provide texts tailored to the individual reader. In particular, the IDAS project (Reiter, Mellish, and Levine 1995) comes closest to the goals of the HealthDoc project in recognizing the need to tailor both textual and non-textual information, including visual formatting, hypertext input, and graphics output. IDAS also emphasizes the need for explicit authoring tools in the document generation process, but here the focus is on authoring at the knowledge-base level, while the HealthDoc authoring tool deals with an actual draft of the document (which can then be translated into a deeper representation). The difference in level of "granularity" of authoring reflects the basic difference in the levels of tailoring done by the two systems: IDAS provides the user with a means of navigating through the whole "hyperspace" of possible texts, but HealthDoc can be seen as providing different variations of the texts at any point in the hyperspace. Consequently, HealthDoc aims to provide a much finer-grained degree of tailoring than does IDAS, which correlates with the difference in their intended types of applications (health information requiring subtle distinctions at all levels of the discourse versus technical documentation needing only three different basic styles).

While IDAS relies mainly on canned texts, other systems use more-dynamic text generation: the Migraine system (Carenini, Mittal, and Moore 1994) uses an approach to text planning that adaptively selects and structures the information to be given to a particular reader. However, Migraine relies on a large number of context-sensitive and user-sensitive text plans so that its methods of tailoring must of necessity be very specific to its particular domain. The PEBA-II system (Milosavljevic and Dale 1996) uses more-general text plans, as well as text templates, that it can choose from to adapt information to the individual reader, but the tailoring done is very specific, focussing on the user's familiarity with a topic. The PIGLET system (Cawsey, Binsted, and Jones 1995) also uses a combination of text plans and text templates, and, like IDAS and Migraine, allows the user to be self-guided in selecting the issues to explore. But its tailoring is also quite specific in nature, mainly concerned with emphasizing material that is relevant to the particular patient. The ILEX-0 system (Knott, Mellish, Oberlander, and O'Donnell 1996) is similar to the PIGLET model in its anticipation of all the possible texts that might be generated, but also includes annotations (e.g., a condition on a piece of canned text) to allow some local customization. However, very free and flexible use of annotations could lead to problems of repetitive text and inappropriate use of referring expressions, the kinds of problems requiring textual repair that HealthDoc's sentence planner is intended to handle.

Like HealthDoc, these systems aim to adapt the style or content of the texts they generate to the characteristics of the individual user, but HealthDoc's approach is more general. HealthDoc allows not only the potential inclusion of explicit text plans and text templates, but a very flexible yet principled means of organizing both textual and non-textual variations, including tailored hypertext, into a 'master document'. Importantly, a HealthDoc, or WebbeDoc, master document can be used by a sentence planner to perform very fine-grained revision and tailoring for a user.

Conclusion

We believe that generation by selection and repair is suitable for applications that exhibit most of the following characteristics:

- **Pre-existing semantic content:** There is a set of variations of a document, or alternative forms, that have been created by a domain specialist. In this case, construction of the master document is greatly facilitated, especially with the use of an authoring tool that supports the assembly and checking of the representations.
- **Similar discourse structure:** The alternative forms all have roughly the same overall structure. This characteristic, which minimizes discourse structure planning or repair, is quite common, especially in cases in which one variation is an expansion of another in level of detail.
- **Granularity at sentence and lexical levels:** Sentences are generally indivisible units, and lexical alternation is limited to simple word and phrase substitutions. These tendencies, which minimize the amount of repair required, help to enforce stylistic uniformity over the range of alternatives.
- **Well-defined criteria for variation:** A clear set of criteria for choosing between alternatives has been identified. Fortunately, in many applications, alternative documents state their intended readership very clearly.

Adaptive-hypertext applications will frequently have these characteristics. Once the core techniques—representation of the master document and methods of repair by sentence planning—are further developed, this model of selection and repair may become one of the most attractive and popular approaches to developing useful Web-based systems that can tailor a document, whether ordinary text or hypertext, to the individual reader.

Acknowledgements

The HealthDoc Project is supported by a grant from Technology Ontario, administered by the Information Technology Research Centre. Vic DiCiccio was instrumental in helping us to obtain the grant, and has been invaluable in subsequent administration. Some material in the section on the functions of sentence planning was written by Daniel Marcu; it is used here with his permission. The other members of the HealthDoc Project have also contributed to the work described here, especially Kim Parsons, Mary Ellen Foster, and Phil Edmonds.

References

Banks, Steven (1997). Master's thesis. Department of Computer Science, University of Waterloo, expected Spring 1997.

Bateman, John Arnold (1995). "KPML: The KOMET–Penman multilingual linguistic resource development environment." *Proceedings, 5th European Workshop in Natural Language Generation*, Leiden, May 1995, 219–222.

Brusilovsky, Peter (1996). "Methods and techniques of adaptive hypermedia." *User Modeling and User-Adapted Interaction*, **6** (2–3), Special Issue on Adaptive Hypertext and Hypermedia, 1996, 87–129.

Carenini, Giuseppe; Mittal, Vibhu O.; and Moore, Johanna D. "Generating patient-specific interactive natural language explanations." *Proceedings, Eighteenth Annual Symposium on Computer Applications in Medical Care*, Washington D.C., November 1994, 5–9.

Cawsey, Alison; Binsted, Kim; and Jones, Ray. "Personalised explanations for patient education." *Proceedings of the Fifth European Workshop on Natural Language Generation*, 1995, 59–74.

DiMarco, Chrysanne and Banks, Steven (1997). "Using subsumption classification on a stylistic hierarchy as the basis of a multi-stage conversion of natural language text to sentence plans." In preparation.

DiMarco, Chrysanne and Foster, Mary Ellen (1997). "The automated generation of Web documents that are tailored to the individual reader." To appear in *Proceedings, 1997 AAAI Spring Symposium on Natural Language Processing for the World Wide Web*, Stanford University, March 1997.

DiMarco, Chrysanne; Hirst, Graeme; Wanner, Leo; and Wilkinson, John (1995). "HealthDoc: Customizing patient information and health education by medical condition and personal characteristics." *Workshop on Artificial Intelligence in Patient Education*, Glasgow, August 1995.

Hovy, Eduard and Wanner, Leo (1996). "Managing sentence planning requirements." *Proceedings, ECAI-96 Workshop on Gaps and Bridges: New Directions in Planning and Natural Language Generation*, Budapest, August 1996.

Knott, Alistair; Mellish, Chris; Oberlander, Jon; and O'Donnell, Mick. "Sources of flexibility in dynamic hypertext generation." *Proceedings, Eighth International Natural Language Generation Workshop*, Herstmonceaux Castle, June 1996, 151–160.

Mann, William C. and Thompson, Sandra A. (1988). "Rhetorical Structure Theory: Toward a functional theory of text organization." *Text*, **8**(3), 1988, 243–281.

Marcu, Daniel (1997). "From local to global coherence: A bottom-up approach to text planning." Submitted for publication.

Milosavljevic, Maria and Dale, Robert. "Strategies for comparison in encyclopaedia descriptions." *Proceedings, Eighth International Natural Language Generation Workshop*, Herstmonceaux Castle, UK, June 1996, 161–170.

Parsons, Kimberley J. (1997). Master's thesis, Department of Computer Science, University of Waterloo, expected Spring 1997.

Penman Natural Language Group (1989). "The Penman primer", "The Penman user guide", and "The Penman reference manual." Information Sciences Institute, University of Southern California.

Reiter, Ehud; Mellish, Chris; and Levine, John. "Automatic generation of technical documentation." *Applied Artificial Intelligence*, **9**, 1995, 259–287.

Wanner, Leo and Hovy, Eduard (1996). "The HealthDoc sentence planner." *Proceedings of the Eighth International Workshop on Natural Language Generation*, Brighton, UK, June 1996.