© *The Journal on Information Technology in Healthcare* 2008; **6**(1): 63–71

Authoring and Generation of Individualised Patient Education Materials

Chrysanne DiMarco[*], *Peter Bray*[†], *Dominic Covvey*[§], *Donald Cowan*[*§], *Vic DiCiccio*[*], *Eduard Hovy*[#], *Joan Lipa*[†], *Cathy Yang*[*]

[*]David R. Cheriton School of Computer Science, University of Waterloo, [†]Division of Plastic Surgery, University of Toronto, [§]Waterloo Institute for Health Informatics Research, Ontario, Canada, and [#]Information Sciences Institute, University of Southern California, USA.

ABSTRACT

E-health services are playing an increasingly important role in healthcare management by providing relevant and timely information to patients about their medical care. An important factor in the rapid growth of online health services is the trend in health management to patient-centric healthcare. An effective means of providing patient-centric healthcare is through the personalisation of health information. Individually tailored information will enable patients to be better educated about their specific condition and also to make informed decisions. To achieve this aim we are developing Natural Language Generation tools and methodologies that can be used to deliver health educational materials tailored to the needs of an individual patient in a timely and accessible manner through Web-based systems. Our initial domain of application is reconstructive breast surgery, but the Natural Language software tools and authoring methodologies we are producing will be generally applicable to all medical interventions.

INTRODUCTION

Present-day health education is often limited in its effectiveness by the need to address it to a wide audience. Options range from producing a minimal, generic document containing only the information that is common to everyone, to a comprehensive document containing all the information that could be relevant to any individual patient. Inevitably the latter approach results in a document that contains much irrelevant information for many patients. This poses a challenging problem as material containing irrelevant information or omitting relevant information can give the reader the impression that it is not addressed at them. As a result they are

Correspondence and reprint requests: Chrysanne Di Marco, PhD, David R. Cheriton School of Computer Science, University of Waterloo, 200 University Avenue West, Waterloo, Ontario, Canada N2L 3G1. E-mail: cdimarco@uwaterloo.ca.

likely to discount or ignore it with consequent problems in compliance with medical regimens.

Studies in health communication have shown that health education material can be much more effective if it is customised for the individual reader in accordance with their specific medical condition, demographic variables, personality profile, and other relevant factors. In one series of studies, unsolicited leaflets were sent to family practice patients on topics such as giving up smoking[1], improving dietary behaviour[2], or having a mammogram[3]. Each leaflet was tailored to the recipient on the basis of data gathered from them in an earlier survey. In each study, the tailored leaflets were found to have a significantly greater effect on patients' behaviour than generic leaflets given to patients in the control group.

To create individualised patient information, researchers have attempted to use Natural Language Generation (NLG) and methods from Artificial Intelligence and Computational Linguistics to develop automated systems for tailoring health information to individual patients[5–10]. In a previous project (HealthDoc) we developed a method for generating tailored documents based on a new paradigm for NLG – 'generation-by-selection-and-repair'. In this technique new documents are created from a pre-existing 'Master Document' containing all the pieces of text that might be needed in tailoring a version of the document for any particular audience. Selections from the master document are made for both content and form, and then are automatically 'repaired' for form, style, and coherence. In this paper we describe some of the challenges and approaches to providing individualised patient education materials, and illustrate these using reconstructive breast surgery. This is a representative example of the kinds of problems that need to be addressed in developing automated Natural Language tailoring systems, as interventions in modern surgical oncology are often complex, multi-step procedures involving multiple surgeons or surgical subspecialties.

TAILORED PATIENT EDUCATION IN RECONSTRUCTIVE SURGERY

Modern reconstructive plastic surgery has evolved into a highly complex field aimed at restoration of patient form and function. The surgical solution to a given reconstructive problem may require grafts of various types (skin, bone, and tendon) combined with tissue-mobilising procedures (flaps) from different locations on the body. Each individual reconstruction will have different implications for aesthetics, function, rehabilitation, recovery, and potential complications, all of which must be reviewed with the patient preoperatively.

In many surgical specialities, brochures, Internet websites, and other forms of 'take-home' educational materials are frequently used to supplement the surgeon-patient consultation and enhance patient retention of information. However, such solutions have proven impractical for much of reconstructive plastic surgery due to the sheer number of techniques available and their frequent need to be performed in

combinations. To address this problem we are trying to develop a system for generating preoperative patient education materials that tailors the text to each individual patient regardless of the complexity of their surgical intervention.

The components of the system are:

- An NLG tailoring system
- Content authoring environment
- Creation of a database of educational modules pertaining to each subcomponent of a given surgical intervention

Our earlier HealthDoc Project demonstrated that complex, stylistically polished texts can be crafted from pre-existing texts represented in an appropriate Master-Document format. In this study we are continuing the development of the 'generation-by-selection and-repair' paradigm, with particular emphasis on designing authoring environments needed to support text-to-text generation systems. Our long-term goal is to develop methods and architectures for Natural Language Generation systems based on selection-and-repair that can be used to produce high-quality, finely tailored documents.

AUTOMATED GENERATION OF TAILORED HEALTH-EDUCATION DOCUMENTS

Our approach is similar in its underlying goal of providing more-effective, patient-centric health information through content geared to the individual. However, our approach to Natural Language Generation relies on the *pre-authoring* of reusable content and a subsequent process of selection, re-assembly, and revision to create tailored versions of a document. The key components in our tailoring system are an *authoring tool* that assists the physician to directly enter textual variants into the HealthDoc Master-Document format, and a 'repair engine' that automatically 'repairs and polishes' the selected text. To develop such a tailoring system, a number of research issues in Natural Language Generation need to be addressed:

- Representation of the Master Document
- Authoring and knowledge-based document management
- 'Sentence planning' methods for automated post-editing

A number of previous Natural Language Generation systems have automatically produced texts tailored to the individual reader.

PAULINE was an early landmark system in the field, setting an example of how formal text-planning methods could be used in creating stylistically varied texts[11]. The development of sophisticated Natural Language systems capable of generating linguistically complex, stylistically varied texts is still an open research issue in this field. Migraine[4], IDAS[7], PIGLET[6], PEBA-II[12], and ILEX[13] are examples of earlier systems that achieved a degree of stylistic control in generation. More recent work on tailoring of Natural Language Generation has involved personalisation of multimedia presentations and stylistic variation of spoken dialogue[14–16]. Many research

issues still remain to be solved in order to build Natural Language Generation systems capable of the kind of fine-grained planning of lexical choice, sentence formation, expression of discourse relations, and text structure needed to produce highly varied, high-quality tailored documents.

AUTHORING TOOLS FOR NATURAL LANGUAGE GENERATION

The creation of the input specifications for Natural Language Generation systems is a problem for all generation systems. A major difficulty is the issue of where the information is obtained that specifies the actual text to be generated[17]. Our approach to this problem is 'preparing' a database of input specifications by authoring material for later use in generating new documents. We believe that this methodology will reduce the complexities inherent in trying to generate high-quality, stylistically expressive text 'from scratch'.

A number of other approaches to natural language authoring have been developed[18–21]. One for example focuses on 'natural language authoring' in contrast to the traditional Natural Language Generation approach in which "*the semantic input is provided interactively by a person rather than by a program accessing digital knowledge presentations*." Our approach to authoring for Natural Language Generation falls somewhere between this paradigm and that of traditional generation systems. As others do with authoring-based document-generation systems, we allow a user to enter the exact textual input that will later be used in creating new texts[22], but we also deal with authoring of input at a deeper level of linguistic representation[23,24], as is typical of traditional generation systems. Up until now, in patient education applications, authoring of materials has typically been accomplished through the interaction of the health professional with a 'knowledge engineer', someone trained in structured knowledge acquisition. Our goal is to provide a supportive authoring environment that will allow the domain experts – physicians, surgeons, and other healthcare providers – to interact directly with the Natural Language tailoring system to enter the textual variants that will later be used to produce the tailored versions.

THE NEED FOR COMPUTER-AIDED AUTHORING IN CREATING PERSONALISED PATIENT EDUCATION MATERIALS

The fundamental problem in personalisation of patient education is that the process involves much more than just producing a brochure or leaflet in half a dozen different versions for different audiences. In practice the number of different combinations of factors can easily be in the tens or hundreds of thousands. It is consequently impossible to produce and distribute, in advance of need, the large number of different versions of each publication necessary for individual tailoring of health information.

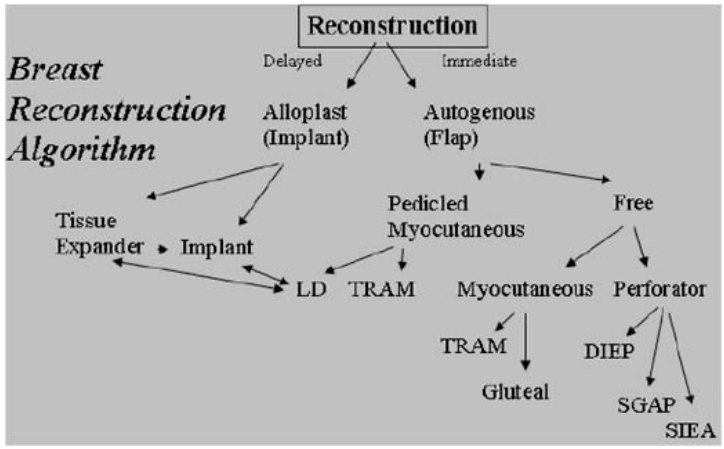

Figure 1. *Decision tree of surgical options in breast reconstruction*

DIEP = Deep Inferior Epigastric Perforator, LD = Latissimus Dorsi, SGAP = Superior Gluteal Artery Perforator Flap, SIEA = Superficial Inferior Epigastric Artery Flap, TRAM = Trans-rectus Abdominis Muscle.

If one just considers reconstruction after radiotherapy, a variety of surgical options may be available for a given situation. Each of these options will be associated with different advantages, disadvantages, and peri-operative implications. It is a challenge for both patients and surgeons to ensure that sufficient information has been communicated preoperatively about these procedures. Moreover, the complexity of the surgical procedure and the variety of options that need to be considered in tailoring documentation to the individual patient makes the creation of appropriate material a potentially explosive process (Figure 1).

To provide appropriate information about the various options, the surgeon needs a database of educational modules pertaining to each subcomponent of a given surgical intervention. Our approach to providing such educational material is based on the *authoring*, in advance of need, of a 'Master Document' library of text fragments which indexes each fragment by a set of features that indicate the conditions under which the fragment should be included in the output. These features represent significant characteristics of the patient, the situation, the desired document style and tone, etc. In order to use our Master-Document representation for breast reconstructive surgery, the phases of a given surgical procedure are broken down into component parts, including: preoperative investigations, procedure description, sequelae, complications, etc. Obviously, authoring and maintaining this type of variable educational material, which must be customised according to treatment options, and potentially a large number of patient characteristics, is an extremely complex process. Our objective is to provide a supportive authoring environment that will guide surgeons and other healthcare providers to directly create customised educational content, without the need for programmer involvement.

THE BASIC AUTHORING TOOL

We have been developing the Authoring Tool in stages, beginning with a simple tool that just displays the Master Document structure. This basic tool uploads an Excel spreadsheet of content variations and displays the Master Document 'discourse tree' hierarchical structure, starting from the topmost node representing the overall document, down through sections, topics, and individual sentences. Each of these document elements is displayed, together with its set of possible variations and Boolean selection condition on each variation. With this simple tool, the intent was to give the surgeon a means of viewing the discourse structure of their manually authored text, often very dense and 'block'-like, in order to better identify where the text could be broken down into smaller, more understandable, segments. The visual display of the document variants also allows the surgeon to review the selection condition attached to each variant, and to check the validity and consistency of the conditions for a set of variants. At this stage though, the authoring tool is 'read-only': the surgeon can view the Master Document but must still return to the Excel spreadsheet to do updates and to preview customised versions using our Web-based Tailoring System.

In the subsequent stage of development, we designed a generic Authoring Tool intended for creation of tailorable Master Documents in any domain, i.e., the tool provides essential features for creating and maintaining tailorable documents, but does not yet provide the advanced features specific to physician needs, e.g., modelling physician 'workflow', adaptability to individual physician practices. At this stage, the Authoring Tool provides the following capabilities:

- Web accessibility to permit use of Authoring Tool independent of location
- Minimal input required from author
- Editing of content variants and selection conditions
- Split-pane interface to allow simultaneous viewing of content, selection conditions, preview of tailored output
- Additional schematic ('discourse tree') view of content
- Ability to preview tailored versions and return to editing content

THE PHYSICIAN'S AUTHORING ENVIRONMENT

The next stage of development will focus on evolving the initial generic tool to an authoring environment geared to the specific needs of physicians, such as adaptability to individual provider, medically 'intelligent' assistance, and knowledge of high-level medical discourse. This will involve providing:

- Capability for provider-specific modifications of individual procedures or content modules
- Both *push* and *pull* methods of content authoring, i.e. the author may directly enter procedure-specific text component (push) or use a wizard-like query engine which intelligently prompts for details (pull)

- Allowance for cut-and-paste or drag-and-drop editing of text elements, such as between similar surgical procedures
- Incorporation of 'Knowledge Level' modelling of domain information

In authoring the content variants, the surgeon will trace through his or her typical physician-patient dialogue associated with the stages in the reconstructive surgical procedure. As a consequence, it is essential that the authoring tool be adaptive to the needs of the individual surgeon, so that the structure of the Master Document may be modified according to the order in which the surgeon wishes to present the information.

Up to this stage, the development of the authoring environment has focused primarily on the 'content level' of patient information, i.e. the actual text to be given to patients. However, there is no 'knowledge-level' modelling for knowledge acquisition to support the creation of tailored educational content. At the knowledge level of authoring, the physician will be guided through the process of considering the concerns of the various stakeholders (e.g. surgeons, patients, hospital) with regard to tailoring the educational material. For example, the surgeon may be primarily concerned with communicating information that will ensure patient compliance with the recommended treatment and that will lead to favourable outcomes whereas the patient may be most concerned with the variations in risks and complications associated with the different treatment options. The authoring tool should therefore ideally embody a cognitive model that aids the physician in mapping out these complementary, and sometimes contradictory, high-level concerns.

The Constructivist theory model can be used as the basis for creating an authoring tool for patient-centred learning and to guide the physician through the process of creating the Master-Document framework. The approach assumes that learners construct their own knowledge from their experiences and that the educator is only the knowledge provider. The Constructivist theory has been previously used to develop a patient education model and to design a knowledge acquisition framework to assist health professionals in organising their domain knowledge prior to the writing of the actual textual content[25].

DISCUSSION

We are developing Natural Language Generation tools and methodologies that can be used to deliver health educational materials tailored to the needs of the individual patient in a timely and accessible manner through Web-based systems. When this project is completed it will provide important software tools and document authoring methodologies to assist in providing patient-centric healthcare. The methodologies for authoring tailorable health education content, from acquiring domain knowledge about the specific medical intervention to writing the actual text variants that will be used in generating tailored patient education documents, will be a significant contribution to advancing research on the uses of tailoring in patient

education, and will be widely applicable in promoting the use of personalisation in health information systems as a means of delivering more effective healthcare.

Our initial domain of application is reconstructive breast surgery, but the Natural Language software tools and authoring methodologies we will develop will be generally applicable to all medical interventions. The methodology will provide a means of shaping complex information so that it is more relevant and personalised. As a consequence it will be able to effectively assist in addressing complex issues such as the achievement of informed consent for surgical procedures. This has been shown to be an important component for improving patient engagement and compliance with medical regimens, and a technique for complementing and reinforcing the information communicated during the pre-surgical encounter. The Natural Language software tools in conjunction with an authoring tool to be used by physicians in creating the tailored content, and a tailoring engine for generating personalised versions for individual patients, will form a robust architecture to allow healthcare providers to expand the educational scope beyond reconstructive surgery to all forms of medical intervention, surgical or otherwise.

ACKNOWLEDGEMENTS

We would like to acknowledge the financial support of Bell University Laboratories and the Natural Sciences and Engineering Research Council of Canada.

REFERENCES

1 Strecher VJ, Kreuter M, Den Boer DJ, Kobrin S, Hospers HJ, Skinner CS. The effects of computer-tailored smoking cessation messages in family practice settings. *The Journal of Family Practice* 1994; **39**: 262–70.

2 Campbell MK, DeVellis BM, Strecher VJ, Ammerman AS, DeVellis RF, Sandler RS. Improving dietary behavior: the effectiveness of tailored messages in primary care settings. *American Journal of Public Health* 1994; **84**: 783-87.

3 Skinner CS, Strecher VJ, Hospers H. Physicians' recommendations for mammography: do tailored messages make a difference? *American Journal of Public Health* 1994; **84**: 43–49.

4 Carenini G, Mittal V, Moore JD. Generating patient-specific interactive natural language explanations. *Proceedings of the Eighteenth Annual Symposium on Computer Applications in Medical Care*, Washington D.C 1994, pp. 5–9.

5 Cawsey AJ, Webber BL, Jones RB. Natural language generation in health care. *Journal of the American Medical Informatics Association* 1997; **4**: 473–82.

6 Cawsey AJ, Binsted K, Jones R. Personalised explanations for patient education. *Proceedings of the Fifth European Workshop on Natural Language Generation*, 1995, pp. 59–74.

7 Reiter E, Mellish C, Levine J. Automatic generation of technical documentation. *Applied Artificial Intelligence* 1995; **9**: 259–87.

8 Reiter E. Osman L. Tailored patient information: Some issues and questions. *Proceedings of the Conference of the Association for Computational Linguistics*, 1997.

9 Bental DS, Cawsey AJ, Jones R. Patient information systems that tailor to the individual. *Patient Education and Counselling* 1999; **36**: 171–80.

10 Green NL. Analysis of linguistic features associated with point of view for generating stylistically appropriate text. *AAAI Spring Symposium: Exploring Attitude and Affect in Text: Theories and Applications*, AAAI Press, 2004.
11 Hovy E. Pragmatics and natural language generation. *Artificial Intelligence* 1990; **43**: 153–97.
12 Milosavljevic M, Dale R. Strategies for comparison in encyclopaedia descriptions. *Proceedings of the Eighth International Natural Language Generation Workshop*, 1996, pp. 161–70.
13 Knott A, Mellish C, Oberlander J, O'Donnell M. Sources of flexibility in dynamic hypertext generation. *Proceedings of the Eighth International Natural Language Generation Workshop*, June 1996, pp. 151–60.
14 Androutsopoulos I, Kokkinaki V, Dimitromanolaki A, Calder J, Oberlander J, Not E. Generating multilingual personalized descriptions of museum exhibits: the M-PIRO project. *Proceedings of the International Conference on Computer Applications and Quantitative Methods in Archaeology*, Gotland, Sweden, 2001.
15 Fleischman M, Hovy E. Towards emotional variation in speech-based natural language processing. *Proceedings of the International Conference on Natural Language Generation*, 2002.
16 Moore JD, Foster ME, Lemon O, White M. Generating tailored, comparative descriptions in spoken dialogue. *Proceedings of the17th Annual FLAIRS Conference*, 2004.
17 O'Donnell M, Knott A, Oberlander J, Mellish C. Optimising text quality in generation from relational databases. *Proceedings of the First International Natural Language Generation Conference*, 2000.
18 Hartley A, Paris C. Multilingual document production: From support for translating to support for authoring. *Machine Translation, Special Issue on New Tools for Human Translators*, 1997.
19 Power R, Scott D. Multilingual authoring using feedback texts. *Proceedings of the 17th International Conference on Computational Linguistics and 36th Annual Meeting of the Association for Computational Linguistics*, 1998.
20 Lapalme G, Brun C, Dymetman M. XML based multilingual authoring. In: V. Keselj and T. Endo (eds.), *Proceedings of the PACLING Conference*, 2003.
21 Brun C, Dymetman M, Lux V. Document structure and multilingual authoring. *Proceedings of the First International NaturalLanguage Generation Conference*, 2000.
22 Parsons. An authoring tool for customizable documents. M.Math. thesis, University of Waterloo, 1997.
23 Banks S. English to sentence plan translation using subsumption on stylistic abstraction. M.Math. thesis, University of Waterloo, 1999.
24 Banks S, DiMarco C. Surface form to sentence plans: A method for English-to-English translation. *Proceedings of the 4thSymposium on Natural Language Processing*, 2000.
25 Yang C. Design of a knowledge acquisition tool using a constructivist approach for creating tailorable patient education materials. M.Math. thesis, University of Waterloo, 2005.