# Automating Personalized Patient Information in Runyankore

**Joan Byamugisha**
University of Cape Town
Cape Town, South Africa
jbyamugisha@cs.uct.ac.za

**Brian DeRenzi**
University of Cape Town
Cape Town, South Africa.
bderenzi@cs.uct.ac.za

**C. Maria Keet**
University of Cape Town
Cape Town, South Africa
mkeet@cs.uct.ac.za



## Abstract
Studies have shown that engaging patients in their treatment, by providing them with information about their medical condition, can lead to better health outcomes. However, generic medical information targeted at a wide audience has been found to be limited in its effectiveness. This has led to efforts to produce personalized patient information. Personalizing information in the healthcare domain is complicated because differences among patient characteristics (age, preexisting conditions, genetic predispositions, etc.) can be in the tens or hundreds of thousands. This has nonetheless been best achieved by Intelligent User Interfaces (IUIs) applying Natural Language Generation (NLG). Most NLG healthcare systems apply templates—where information is entered into predetermined slots—as a means of generating customized patient information. Templates are inapplicable to Bantu languages, due to their characteristic agglutinative structure. We present here our ongoing efforts to develop an NLG system for Runyankore, a Bantu language indigenous to Uganda, with the aim of using it to produce personalized patient information in rural Uganda.

## Author Keywords
NLG; Runyankore; IUI, grammar engine, ontology verbalization, personalized health information

## ACM Classification Keywords

I.2.7 [Natural language Processing]: Language Generation

## Introduction

There is a new trend towards patient-centric healthcare, which aims to involve patients directly in the medical decision-making process, through providing them with the relevant information they need to better understand their medical condition [6, 11, 28]. This leads to patients making more informed decisions about their prescribed treatment [10, 13], which in turn leads to better patient outcomes and reduced healthcare costs [6, 10, 13, 16, 28]. The provided patient information can be used to complement and reinforce what is discussed during the patient-doctor consultation [6, 8, 11], especially given that the fraction of the information which is actually retained by the patient is consistently rather small [8, 11, 9].

The problem with most patient-information material is its limited effectiveness when generalized to apply to a wide audience [8, 11, 12]. What is usually produced is either a generic document with minimal information common to everyone [6, 11, 28], or a large document which tries to provide the maximum information considered relevant to one (and hence mostly irrelevant to many) [8, 11, 9, 12]. Such material is likely to be discounted or ignored by patients [8, 9, 12]. On the other hand, studies in health communication have shown that patient information is likely to be more effective if it is personalized for a specific patient [6] and presented in an understandable form and manner [13], that is, natural language.

Intelligent User Interfaces (IUIs) focused on automated output generation of relevant patient content have been achieved through Natural language Generation. NLG, being the production of understandable texts in a selected human language from an underlying non-linguistic representation of information [2, 25], has been successfully applied to produce personalized patient information [6, 12, 7, 18, 22, 23] using templates. Templates are selected predefined structures, with blank spaces that usually have associated requirements specifying what kind of information can fill them [17, 25]. Our research aims to extend the advantages of personalized patient information to rural Uganda by automatically generating text in Runyankore.

Runyankore is a Bantu language indigenous to south western Uganda [1, 26, 27]. It has the characteristic complex agglutinative structure (glueing together of different grammatical units to form a single word), noun class system, and verbal morphology [1, 26, 27]. This makes the use of templates in the generation of text in Runyankore inapplicable. We discuss here our use of another NLG technique, grammar engine, to capture the complex linguistic structure of Runyankore, and generate text using ontologies as input into our NLG system. An ontology is a logical theory specifying the entities of interest in a given domain, and the relations and constraints which hold among these entities [14, 15]. The process of generating natural language descriptions from these logical theories is called verbalization.

## Automated Content Generation in Runyankore

Our research is aimed at automating the generation of patient information in Runyankore, particularly drug prescription explanations. Our interest in IUIs as human-computer interfaces lies in their ability to improve the efficiency, effectiveness, and naturalness of HCI [24]. Our IUI implementation uses semantic web technologies, and verbalizes knowledge represented in the Web Ontology Language (OWL) in particular.

Content generation has been achieved by using the NLG grammar engine technique. A grammar engine considers grammatical categories (such as sentence, noun phrase, and verb phrase) and rules which implement the categories as objects with complex sets of properties associated with them [20]. In our case, the grammatical categories are: subject prefix, adjective prefix, genitive, grammatical form, tense, and aspect. The first four are determined by the noun class of the subject or object, while the last two are necessary for verb conjugation. The rules generally include: patterns of interpreting logical theories [3] based on work done for isiZulu [21], obtaining plurals of nouns [5], verb conjugation [4], and phonological conditioning.

The input into the NLG system is a logical theory, and the output a textual interpretation of that logical theory in Runyankore. The example shows the verbalization of subsumption ($\sqsubseteq$), a relation which states that the entity to the left of $\sqsubseteq$ is a subclass of the entity to the right (the verbalization of the constructor itself is bolded):

**Logical theory:** Hydrocodone $\sqsubseteq$ MorphineDerivative

**English:** **Every** hydrocodone **is a** morphine derivative

**Runyankore:** **Buri** mubazi gwa hydrocodone **n'**omubazi gw'okukyendeeza obusaasi

Our NLG system has been tested on a sample of a large healthcare ontology—SNOMed-CT [19]—in order to assess the generation of Runyankore text containing medical jargon. Our IUI is able to select content based on the logical theories of interest in the ontology. The generated text is then written to a text file.

## Evaluation of Generated Text

The typical method of evaluating the performance of NLG systems is to ask subjects to read and judge the generated text, as compared to human-authored text [2] We are taking a qualitative community-based approach to IUI evaluation [24], by testing naturalness in two ways: (1) assessing the grammatical correctness and/or understandability, and (2) distinguishing between human-authored and computer generated text. As our NLG system is based on a grammar engine, it is crucial to test its output for its adherence to Runyankore grammar rules. We plan to present study participants with different sentences, which they will then categorize based on their assessment of the correctness of the grammar, and how well they understand each. Distinguishing between human-authored and generated text will be done by presenting study participants with a list of sentences, composed of both human-authored and generated text. They will then label each sentence, based on whether they assess it to have been written by a human or generated by a computer.

Our target population is all people who can read, write, and speak Runyankore. This is subdivided into two: linguists and non-linguists. Runyankore Linguists are important in this study because they have a deep understanding of the grammar of the language, and can thus provide insightful feedback about what aspects of the generated text are incorrect and/or not understandable. Non-linguists, on the other hand, are the audience to which the applications of our NLG system will be put. It is therefore important to ensure that they understand the generated text, and consider it to be correct. We are considering obtaining our study participants from Mbarara, a district in Uganda where Runyankore is predominantly and ethnically spoken.

## Conclusion and Future Work

The benefits of providing personalized patient information have led to efforts to automate the generation of text in the healthcare domain. This has been achieved through IUIs based on NLG systems. Our ongoing efforts to implement similar systems for Runyankore have resulted in an NLG system based on a grammar engine. We are currently evaluating the generated text, which may lead to a revision of the NLG system, to account for the recommendations given. In this case, the updated NLG system will be re-evaluated with new study participants. Once the NLG system is completed, the targeted initial application is to generate personalized drug prescription explanations.

## References


[1] Allen Asiimwe. 2014. *Definiteness and Specificity in Runyankore-Rukiga*. Ph.D. Dissertation. Stallenbosch University, Cape Town, South Africa.

[2] Nadjet Bouayad-Agha, Gerard Casamayor, and Leo Wanner. 2012. Natural Language Generation in the Context of the Semantic Web. *Semantic Web Journal* (2012).

[3] Joan Byamugisha, C. Maria Keet, and Brian DeRenzi. 2016a. Bootstrapping a Runyankore CNL from an isiZulu CNL. In *CNL 2016*. Springer LLNCS, Aberdeen, Scotland.

[4] Joan Byamugisha, C. Maria Keet, and Brian DeRenzi. 2016b. Tense and Aspect in Runyankore using a Context-Free Grammar. In *INLG 2016*. Edinburgh, Scotland.

[5] Joan Byamugisha, C. Maria Keet, and Langa Khumalo. 2016c. Pluralizing Nouns in isiZulu and Related Languages. In *CICLing 2016*. Konya, Turkey.

[6] J. Alison Cawsey, B. Ray Jones, and Janne Pearson. 2000. The Evaluation of a Personalized Health Information System for Patients with Cancer. *User Modeling and User-Adapted Interaction* 10, 1 (2000), 47–72.

[7] Fiorella de Rosis, Floriana Grasso, and C. Dianne Berry. 1999. Refining Instructional Text Generation after Evaluation. *Artificial Intelligence in Medicine* 17, 1 (1999), 1–36.

[8] Chrysanne DiMarco, Peter Bray, Dominic Covvey, Don Cowan, Vic DiCiccio, Eduard Hovy, Joan Lipa, and Cathy Yang. 2005. Authoring and Generation of Tailored Preoperative Patient Education Materials. In *Workshop on Personalization in e-Health User Modeling, Conference*. Edinburgh, Scotland.

[9] Chrysanne DiMarco, Peter Bray, Dominic Covvey, Don Cowan, Vic DiCiccio, Eduard Hovy, Joan Lipa, and Cathy Yang. 2006. Authoring and Generation of Individualized Patient Education Materials. In *Conference of the American Medical Informatics Association*. Washington D. C.

[10] Chrysanne DiMarco, Dominic Covvey, Peter Bray, Don Cowan, Vic DiCiccio, Eduard Hovy, Joan Lipa, and Doug Mulholland. 2007. The Development of a Natural Language Generation System for Personalized e-Health Information. In *Medinfo 2007*. Brisbane, Australia.

[11] Chrysanne DiMarco, Don Cowan, Peter Bray, Dominic Covvey, Vic DiCiccio, Eduard Hovy, Joan Lipa, and Doug Mulholland. 2006. A Physician's Authoring Tool for Generation of Personalized Health Education in Reconstructive Surgery. In *AAAI Spring Symposium on Argumentation for Consumers of Healthcare*. Stanford University.

[12] Chrysanne DiMarco, Graeme Hirst, Leo Wanner, and John Wilkinson. 1995. HealthDoc: Customizing Patient Information and Health Education by Medical Condition and Personal Characteristics. In *Workshop on Artificial Intelligence in Patient Education*. Glasgow, Scotland.

[13] Chrysanne DiMarco, David Wiljer, and Eduard Hovy. 2009. Self-Managed Access to Personalized Healthcare through Automated Generation of Tailored Health Educational Materials from Electronic Health Records. In *AAAI Fall Symposium on Virtual Health Interaction*. Washington D. C.

[14] Scott Farrar and D. Terence Langendoen. 2009. An OWL-DL Implementation of GOLD: An Ontology for the Semantic Web. In *Linguistic Modeling of Information and Markup Languages*. Springer, Chapter 6.

[15] Nicola Guarino, Daniel Oberle, and Steffen Staab. 2009. What is an Ontology? In *Handbook on Ontologies*. Springer, Chapter 6, 1–17.

[16] Graeme Hirst, Chrysanne DiMarco, Eduard Hovy, and Kimberley Parsons. 1997. Authoring and Generating Health-Education Documents that are Tailored to the Needs of the Individual Patient. In *UM 97*. Springer Wien New York, Sardinia, Italy, 107–118.

[17] Eduard Hovy. 1998. Natural Language Generation. *MITECS* (1998).

[18] Sazzad Mohammed Hussain, A. Rafael Calvo, Louise Ellis, Juchen Li, Laura Ospina-Pinillos, Tracey Davenport, and Ian Hickie. 2015. NLG-Based Moderator Response Generator to Support Mental Health. In *CHI'2015*. ACM, Seoul, South Korea.

[19] International Health Terminology Standards Development Organization (IHTSDO). 2015. Systematized Nomenclature of Medicine–Clinical Termes (SNOMED CT). (2015). http://browser.ihtsdotools.org/.

[20] Daniel Jurafsky and H. James Martin. 2007. *Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition*. Prentice Hall, Inc., USA.

[21] C. Maria Keet and Langa Khumalo. 2017. Towards a Knowledge-to-Text Controlled Natural Language of isiZulu. *Language Resources and Evaluation* 51 (2017), 131–157.

[22] Fredrik Lindahl. 2005. Practical Text Generation in Clinical Medicine. (2005). Chalmers Computer Science Department Winter Meeting.

[23] Saad Mahamood and Ehud Reiter. 2011. Generating Affective Natural Language for Parents of Neonatal Infants. In *ENLG 2011*. ACM, Stroudsburg, Pa, USA, 12–21.

[24] Mark Maybury. 1999. Intelligent User Interfaces: An Introduction. In *4th International Conference on Intelligent User Interfaces*. ACM, New York, Los Angeles, California, USA, 3–4.

[25] Ehud Reiter and Robert Dale. 1997. Building Applied Natural Language Generation Systems. *Journal of Natural Language Engineering* 3, 01 (1997), 57–87.

[26] Doreen Daphine Tayebwa. 2014. *Demonstrative Determiners in Runyankore-Rukiga*. Master's thesis. Norwegian University of Science and Technology, Norway.

[27] Justus Turamyomwe. 2011. *Tense and Aspect in Runyankore-Rukiga: Linguistic Resources and Analysis*. Master's thesis. Norwegian University of Science and Technology, Norway.

[28] Lauren Wilcox, Dan Morris, Desney Tan, Justin Gatewood, and Eric Horvitz. 2011. Characterising Patient-Friendly Micro-Explanations of Medical Events. In *CHI'11*. ACM, New York, 29–32.