Natural Language Generation and its advancements

Smith Patterson
Natural Language Processing Lab
Centre for Advanced Technologies
smithp721@gmail.com

Erica Morgan
Natural Language Processing Lab
Centre for Advanced Technologies
ericamorg@gmail.com

Abstract

Natural Language Generation has advanced the way humans are interacted with systems. Also providing different startegies to invlove with systems that are targeting on various complex processes. This paper analyses a set of NLG components with recent developments and also provide latest advaces in the area of Natural Language Generation. Furthermore, we attempt to investigate what new models have arisen in the area of Natural Language Generation.

1 Introduction

Researchers have explored several methods of building NLG systems; modern systems largely employ canned text, features , or templates . In general, it can be said that the document (or text) planning tasks of stage 1 are language-independent but domain-specific, whereas linguistic realisation (stage 3) is language-specific, but can in principle be done in a domain- independent fashion. This difference in domain-dependence of the different tasks (0) largely explains why reusable software packages exist for linguistic realisation, but not for document planning. For example, a template could have the utterance [flight] is departing from [place] at [time]. where [flight], [place], and [time], are to be filled in with canned text. There is some disagreement as to which method(s) are best. In domains where generalization is necessary (e.g., tutoring systems and complex role playing games), using canned text does not scale well. We are primarily interested in template-based methods, as they are well used in practice and involve a shallow learning curve. Recently, researchers have introduced recursively substituted templates (filling in a blank may yield 170 Proceedings of the Seventh AAAI Conference on Artificial Intelligence and Interactive Digital Entertainment Glados Glados Chell Chell Tie a tie Store the store Inform OnSale is on sale at . Sentence Person Item Place Glados (0) says to Chell Figure 1: Example t-template utterance production. Shapes in the top-left of a box denote return types.

2 Closer look to NLG

While feature-based NLG systems can generalize to large domains, they are typically less efficient than other methods; further, effective usage often calls for designers to have specialized linguistic knowledge. Template based approaches have been adopted in many practical systems, but require authors to create template libraries, which becomes more challenging as systems scale. Finally, in situations where speaker and hearer share some visual domain, information on what is mutually visible may be used to produce deictic expressions such as here, there, this and that.

Shapes within an utterance denote a sample parameter configuration. more blanks), blurring what used to be a fine line between template and feature-based approaches . This approach allows templates with similar structure to be more modular and reusable. We utilize typed templates, a derivation of recursively substituted templates that allow templates to have several utterances and restrict templates that can be substituted in blanks. We show that these additional specifications, which can be implemented on top of existing template-based systems, can yield significant improvements in varied natural language as well as reduced language authoring times. In feature-based approaches of NLG, authors specify various parts of speech and lexical items to the NLG system, which in turn synthesizes the information into natural language (0). In template-based approaches, authors use language structures call templates that produce utterances through a fill-in-

theblank mechanism, to realize text.

The tasks associated with stage 2, that of sentence planning, require both domain and language-specic knowledge. For instance, the generation of referring expressions requires knowledge about the application domain (information about domain objects and their properties) and about the application language (0) (e.g., the syntactic properties of modiers expressing certain properties). Almost all NLG systems discussed in the literature are aimed at the gen- eration of stretches of text to be read (or listened to) without interruption by the user (0). NLG is rarely used for the generation of system utterances within a dialogue system; this task is generally performed in a non-linguistic, ad-hoc fashion, using canned text or simple string manipulation.1

Some reasons for this are suggested in the next section, which discusses the dierences between language generation in a 'monologue' or a dialogue situation (0) (0). Consisting of purely hard-coded text, canned text is typically utilized for simple NLG purposes; when prompted for an utterance, the system returns one of the predefined statements.

This may be a problem in application domains involving more sophisticated dialogues where it is less acceptable for the system to behave in a machine-like fashion, such as tutoring and social dialogues. In sum, in many practical dialogue systems there appears to be no direct need for a full-edged language generation component, since relatively simple techniques are sucient to produce adequate language output. However, output generation by means of such simple methods is seriously lacking in exibility and context-sensitivity.

3 System Developments

Speakers also have a tendency to use the same wording as their dialogue part- ner (the 'priming' eect noted by, among others, Levelt and Kelter [1982] and Clark and Wilkes-Gibbs [1986]) (0). In addition, the style of the dialogue utterances (0) depends on the speaker's relation to the listener (e.g., salesperson versus cus- tomer, teacher versus student), on dierent characteristics of the listener (0) (0) (expert or novice, adult or child), and on the speaker's emotional and cognitive state (happy or sad, certain or uncertain).

Ideally, a system for the automatic generation of dialogue (0) utterances should also take these contextual aspects into account. Clearly, doing so using a simple string manipulation approach will be dicult (if not impossible), so a more advanced form of language generation is needed. Human speakers in a dialogue take different (0) aspects of the dialogue context into account when formulating their utterances. For instance, information on the form and content of preceding (0) dialogue utterances is used to produce ellip- tic utterances and pronominalised references to previously mentioned entities.

Different approaches to language generation can be distinguished, and these are discussed in the next section, paying special attention to their usability for the generation of dialogue utterances.

4 Conclusion

Finally, it can be seen that Natural Language Generation has advanced the area of Artificial Intelligence in many ways providing useful functionalities. It is also important to look back to various other types of developments which closely resemble the Natural Language Generation in other different areas of the Computer Science.

References

Eva Banik, Eric Kow, Nikhil Dinesh, Vinay Chaudri, and Umangi Oza. Natural Language Generation for a Smart Biology Textbook. In *INLG 2012 Proceedings of the Seventh International Natural Language Generation Conference*, pages 125–127, Utica, IL, May 2012. Association for Computational Linguistics.

Kevin Bartz, Cory Barr, and Adil Aijaz. Natural language generation for sponsored-search advertisements. In *Proceedings of the 9th ACM conference on Electronic commerce - EC '08*, page 1, New York, New York, USA, July 2008. ACM Press.

Regina Barzilay and Mirella Lapata. Aggregation via set partitioning for natural language generation. In *Proceedings of the main conference on Human Language Technology Conference of the North American Chapter of the Association of Computational Linguistics* -, pages 359–366, Morristown, NJ, USA, June 2006. Association for Computational Linguistics.

Regina Barzilay and Mirella Lapata. Collective content selection for concept-to-text generation. In *Proceedings of the conference on Human Language Technology and Empirical Methods in Natural Language Processing - HLT '05*, pages 331–338, Morristown, NJ, USA, October 2005. Association for Computational Linguistics.

Robert Dale, Donia Scott, and Barbara Di Eugenio. Introduction to the special issue on natural language generation. *Computational Linguistics*, 24(3):346–353, September 1998.

Roger Evans, David Weir, John Carroll, Daniel Paiva, and Anja Belz. Modelling control in generation. In *Proceedings of the Eleventh European Workshop on Natural Language Generation*, pages 25–32, Saarbrücken, Germany, June 2007. DFKI GmbH.

Barbara Di Eugenio, Davide Fossati, Dan Yu, Susan Haller, and Michael Glass. Aggregation improves learning. In *Proceedings of the 43rd Annual Meeting on Association for Computational Linguistics - ACL '05*, pages 50–57, Morristown, NJ, USA, June 2005. Association for Computational Linguistics.

Raquel Hervás and Pablo Gervás. *Progress in Artificial Intelligence*, volume 3808 of *Lecture Notes in Computer Science*. Springer Berlin Heidelberg, Berlin, Heidelberg, December 2005.

Raquel Hervás, Virginia Francisco, and Pablo Gervás. Assessing the influence of personal preferences on the choice of vocabulary for natural language generation. *Information Processing & Management*, 49(4):817–832, July 2013.

Mary Dee Harris. Building a large-scale commercial NLG system for an EMR. pages 157–160, June 2008.

Janusz Kacprzyk and Sawomir Zadrozny. Computing With Words Is an Implementable Paradigm: Fuzzy Queries, Linguistic Data Summaries, and Natural-Language Generation. *IEEE Transactions on Fuzzy Systems*, 18(3):461–472, June 2010.

Paul S. Jacobs. Knowledge structures for natural language generation. In *Proceedings of the 11th coference on Computational linguistics* -, page 554, Morristown, NJ, USA, August 1986. Association for Computational Linguistics.

Kentaro Inui, Takenobu Tokunaga, and Hozumi Tanaka. Text revision: A model and its implementation. In *Sixth International Workshop on Natural Language Generation: Aspects of Automated Natural Language Generation*, pages 215–230, Trente, Italy, 1992. Springer-Verlag.

Kentaro Inui, Atsushi Fujita, Tetsuro Takahashi, Ryu Iida, and Tomoya Iwakura. Text Simplification for Reading Assistance: A Project Note. In *Second International Workshop on Paraphrasing*, pages 9–16, Sapporo, Japan, 2003. Association for Computational Linguistics.

Kathleen R. McKeown. *Text generation using discourse strategies and focus constraints to generate natural language text*. Cambridge University Press, Cambridge, 1985.

David D. McDonald and James D. Pustejovsky. A computational theory of prose style for natural language generation. In *Proceedings of the second conference on European chapter of the Association for Computational Linguistics* -, pages 187–193, Morristown, NJ, USA, March 1985. Association for Computational Linguistics.

Kathleen F. McCoy. Natural language generation and assistive technologies. page 1, May 2012.

Ehud Reiter and Robert Dale. *Building natural language generation systems*. Cambridge University Press, January 2000.

Ehud Reiter and Anja Belz. An Investigation into the Validity of Some Metrics for Automatically Evaluating Natural Language Generation Systems. *Computational Linguistics*, 35(4):529–558, December 2009.

Daniel S. Paiva and Roger Evans. Empirically-based control of natural language generation. In *Proceedings of the 43rd Annual Meeting on Association for Computational Linguistics - ACL '05*, pages 58–65, Morristown, NJ, USA, June 2005. Association for Computational Linguistics.

Rivindu Perera and Parma Nand. The role of linked data in content selection. In Duc-Nghia Pham and Seong-Bae Park, editors, *PRICAI 2014: Trends in Artificial Intelligence*, volume 8862 of *Lecture Notes in Computer Science*, pages 573–586. Springer International Publishing, 2014.

Shimei Pan, Kathleen McKeown, and Julia Hirschberg. Exploring features from natural language generation for prosody modeling. *Computer Speech & Language*, 16(3-4):457–490, July 2002.

Shimei Pan and James C. Shaw. Instance-based sentence boundary determination by optimization for natural language generation. In *Proceedings of the 43rd Annual Meeting on Association for Computational Linguistics - ACL '05*, pages 565–572, Morristown, NJ, USA, June 2005. Association for Computational Linguistics.

Kishore Papineni, Salim Roukos, Todd Ward, and Wei-Jing Zhu. BLEU: a method for automatic evaluation of machine translation. In *Proceedings of the 40th Annual Meeting on Association for Computational Linguistics - ACL '02*, pages 311–318, Morristown, NJ, USA, July 2001. Association for Computational Linguistics.

T. Patten and D.S. Stoops. Real-time generation of natural language. *IEEE Expert*, 6(5):15–22, October 1991.

W.A. Perkins. Generation of natural language from information in a frame structure. *Data & Knowledge Engineering*, 4(2):101–114, August 1989.

Fred Popowich, Milan Mosny, and David Lindberg. Interactive natural language query construction for report generation. pages 115–119, May 2012.

François Portet, Ehud Reiter, Albert Gatt, Jim Hunter, Somayajulu Sripada, Yvonne Freer, and Cindy Sykes. Automatic generation of textual summaries from neonatal intensive care data. *Artificial Intelligence*, 173(7-8):789–816, May 2009.

Owen Rambow. Corpus-based methods in natural language generation. In *Proceedings of the 8th European workshop on Natural Language Generation - EWNLG '01*, volume 8, pages 1–2, Morristown, NJ, USA, July 2001. Association for Computational Linguistics.

Owen C Rambow, Monica Rogati, and Marilyn A Walker. Evaluating a Trainable Sentence Planner for a Spoken Dialogue Travel System. In *Annual Meeting of the Association for Computational Linguistics*, pages 426–433, Sofia, Bulgaria, 2001. Association for Computational Linguistics.

Owen Rambow, Srinivas Bangalore, and Marilyn Walker. Natural language generation in dialog systems. In *Proceedings of the first international conference on Human language technology research - HLT '01*, pages 1–4, Morristown, NJ, USA, March 2001. Association for Computational Linguistics.

Adwait Ratnaparkhi. Trainable approaches to surface natural language generation and their application to conversational dialog systems. *Computer Speech & Language*, 16(3-4):435–455, July 2002.

Ehud Reiter. An Architecture for Data-to-Text Systems. In *Proceedings of the Eleventh European Workshop on Natural Language Generation*, pages 97–104, Saarbrücken, Germany, June 2007. DFKI GmbH.

Ehud Reiter, Sripada Sripada, Jim Hunter, Jin Yu, and Ian Davy. Choosing words in computer-generated weather forecasts. *Artificial Intelligence*, 167(2):137–169, September 2005.

Yuan Ren, Kees Van Deemter, and Jeff Z Pan. Generating Referring Expressions with OWL2. In *Twenty-Third International Workshop on Description Logics*, Waterloo, Canada, 2010. CEUR-WS.

Verena Rieser and Oliver Lemon. Natural language generation as planning under uncertainty for spoken dialogue systems. pages 683–691, March 2009.

Jacques Robin. A Revision-Based Generation Architecture for Reporting Facts in their Historical Context. In *New Concepts in Natural Language Generation: Planning, Realization and Systems. Frances Pinter, London and*, pages 238–265. Pinter Publishers, 1993.

Cesare Rocchi and Massimo Zancanaro. Generation of Video Documentaries from Discourse Structures. In *Ninth European workshop on natural language generation*, Budapest, Hungary, 2003. Association for Computational Linguistics.