Appropriate Microplanning Choices for Low-Skilled Readers

Conference or Workshop Item

Appropriate Microplanning Choices for Low-Skilled Readers

Sandra Williams and Ehud Reiter
Department of Computing Science,
University of Aberdeen, Aberdeen AB24 3UE, U.K.
{swilliam,ereiter}@csd.abdn.ac.uk

Abstract

We have developed a set of microplanning choice rules which are intended to enable Natural Language Generation (NLG) systems to generate appropriate texts for readers with below-average literacy, focusing in particular on choices related to how discourse structure is expressed (cue phrases, ordering, sentence structure). Evaluation experiments suggest that our rules do enhance the readability of texts for low-skilled readers, although there is still room for improvement.

1 Introduction

Natural Language Generation (NLG) systems are computer programs that generate written texts in English or other human languages [Reiter and Dale, 2000]. Most existing NLG systems assume that generated texts will be read by proficient readers with good literacy levels. However, many people in the UK and elsewhere are not proficient readers. The goal of our research is to generate texts which low-skill readers will find (relatively) easy to read.

Generating appropriate texts for poor readers is a multi-faceted problem, which involves choices at all NLG levels (content, microplanning, realisation). The focus of our research is on microplanning choices, in particular on choices related to the expression of discourse structure.

This work was done in the context of the GIRL and Skill-Sum[1] projects. These projects worked in the application area of generating feedback reports on assessments of adult basic skills (e.g. literacy). That is, users took a test assessing their basic skills, and GIRL/SkillSum generated for them reports that summarized their performance on the test.

While many previous researchers have looked at tailoring generated texts according to the user's domain expertise (e.g. [Paris, 1988]), less has been done on tailoring texts according to the reader's literacy. Perhaps the best known previous work in this area is PSET [Devlin *et al*., 1999], which examined choices in texts intended for aphasic readers. Unfortunately most of PSET's rules were not experimentally validated. Scott and de Souza [1990] suggested some psycholinguistically-motivated rules for expressing discourse relations, but did not evaluate them at all.

2 Microplanning choices investigated

The document (content) planners of our systems produce as output a tree, where core messages are related by discourse relations such as ***explanation*** or ***concession***. Discourse relations are essentially RST relations, and messages are represented using a deep-syntactic representation. An example of an extract from a typical content plan, with messages shown as text glosses instead of deep syntactic structures, is shown in Figure 1.

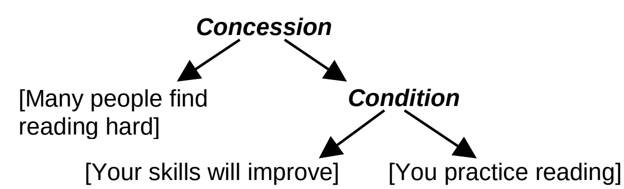

Figure 1 – Extract from a typical content plan

Our work focuses on how discourse relations such as ***Concession*** in Figure 1 are expressed, in particular:

- *cue phrases*: which cue phrases (if any) should be used to express a discourse relation?
- *ordering*: which order should the constituents related by a discourse relation be expressed in?
- *punctuation (sentence structure)*: should constituents be expressed in separate sentences (paragraphs?). If not, should punctuation be used to separate them?

We developed a set of rules for these choices which we hypothesised were appropriate for low-skill readers; this is our Enhanced Readability (ER) model. We also developed a control model for making these choices, based on the most common choices in the RST-DTC [Carlson *et al.*, 2002].

We created a microplanner that generated texts according to the rules in the ER and control models. We used a constraint-based approach that in general terms is similar to Power [2000]; further details are given in Williams [2004].

[1] Funded by PACCIT-LINK Grant ESRC RES-328-25-0026.

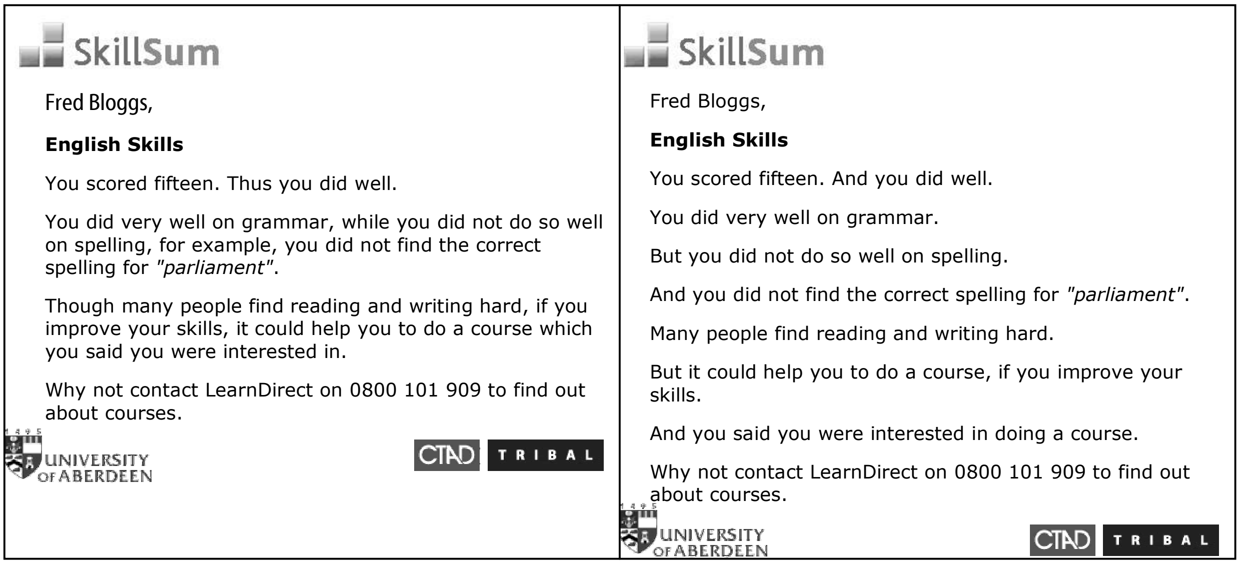

Figure 2: Example texts produced by SkillSum. Left-hand text generated with control model, right-hand text with ER model.

3 Evaluation

GIRL evaluation experiment: 38 subjects, of varied literacy levels, took the GIRL assessment and were shown GIRL's feedback reports. Subjects were initially shown texts generated with either the ER or control version (randomised), and asked to read the report aloud; we measured reading time and reading errors (reading aloud was preferred over silent reading because pilot tests showed that many poor readers would skim texts when asked to silently read). Subjects were then shown both versions (ER and control) of the report, and asked to state a preference.

The results showed that poor readers on average seemed to read ER texts faster, make fewer reading errors on ER texts, and also preferred ER texts; however none of these results were statistically significant. There was little difference between the two versions for good readers

SkillSum evaluation experiment: 60 subjects were selected by skills experts to be people with moderate but not severe literacy problems, we also removed outliers; hence this group was more homogenous that of the first experiment. After completing the assessment and reading their own report, each subject was asked to read a report generated for someone else (in order to de-personalise the experiment); half read ER and half read control versions. In fact the reports read were those shown in Figure 2. As in the GIRL experiment, we measured reading aloud rate and reading errors (but not preference).

This time our results showed a significant effect on reading rate, subjects read the ER version 9% faster than the control version (p=0.04). There was also a weakly significant (p=0.058) improvement in reading errors.

4 Conclusion

We have only scratched the surface of the topic of generating appropriate texts for low-skilled readers; much more can and should be done. In particular we would like to include lexical choice in our models, and also develop different models for people with different skill profiles. Nevertheless, we think our results to date are encouraging, and suggest that good choice rules can make a difference.

References

[Carlson *et al*, 2002] L. Carlson, D. Marcu, and M. E. Okurowski. Building a Discourse-Tagged Corpus in the Framework of Rhetorical Structure Theory. In van Kuppevelt and Smith (eds.), *Current Directions in Discourse and Dialogue*. Kluwer.

[Devlin *et al*, 1999] S. Devlin, J. Tait, Y. Canning, J. Carroll, G. Minnen and D. Pearce. The Application of Assistive Technology in Facilitating the Comprehension of Newspaper Text by Aphasic People. In Buhler and Knops (eds.) *Assistive Technology on the Threshold of the New Millennium*. IOS Press.

[Paris, 1988] C. Paris. Tailoring Object Descriptions to a User's Level of Expertise. *Computational Linguistics* 14(3):64-78.

[Power, 2000] R. Power: Planning texts by constraint satisfaction. *Proc of COLING 2000*, 642-648

[Reiter and Dale, 2000]. E. Reiter and R. Dale. *Building Natural Language Generation Systems*. Cambridge University Press.

[Scott and de Souza, 1990] D. Scott and C. de Souza. Getting the Message Across in RST-Based Text Generation. In R Dale, C. Mellish and M. Zock (eds.), *Current Research in Natural Language Generation*. Academic Press.

[Williams, 2004] S. Williams. Natural language generation (NLG) of discourse relations for different reading levels. PhD Thesis, University of Aberdeen.