Natural language generation of discourse connectives for different reading levels

Sandra Williams
Department of Computing Science,
University of Aberdeen,
Aberdeen AB24 3UE,
swilliam@csd.abdn.ac.uk

Abstract

This paper describes work in progress on the project: Natural Language Generation (NLG) for different reading levels. The project investigates how discourse features affect reading ease. Insights from this investigation will be implemented in an NLG system as one step towards enabling it to generate documents for different reading levels. We describe our test-bed application for experimenting with the NLG system. We also describe a pilot experiment designed to find out whether discourse connectives help, or hinder, adult readers with varying levels of ability.

Background

Seven million adults above the age of 16 years in the UK have some difficulty with reading (Moser, 1999). However, very few adults cannot read at all. The ability to generate documents appropriate for different reading levels will aid low-skilled readers and promote better comprehension at all levels.

At present, NLG systems do not have the ability to tailor documents for different reading levels. We do not know enough about how humans do this. There is a multitude of informal guides, but there has been little real scientific study. Neither do we yet know technically how it can be accomplished by a computer system. In this project we intend to combine knowledge from the fields of basic literacy education, discourse linguistics, psycholinguistics and computer science within our NLG system in order to achieve this goal.

Tailoring for different reading abilities involves making modifications at all linguistic levels, including the lexical level, the syntactic level, the semantic level and the discourse level. An NLG system will need make appropriate choices at each level. This raises issues about whether such choices will change the meaning and, if so, then how much will it change? Will some information be lost?, and, if so, how much information loss is acceptable? At the lexical level, for instance, a common short word could be selected rather than an uncommon long word, but there is no exact equivalence in meaning for two different words. Indeed, people may not even agree on the meaning of a single word. Because of the wide range of factors involved at every linguistic level, we have decided to concentrate this study on discourse level features. First, we are considering the effects of discourse connectives on reading ease. Discourse connectives make explicit the rhetorical relationships between different parts of a document. Therefore our study will necessarily include rhetorical structure and will overlap with other structural considerations such as layout and information structure.

One of our main sources of information about what adult literacy learners might be able to read is the UK Adult Literacy Core Curriculum (ALCC) (Steeds, 2001). This curriculum identifies five levels of adult learners. There are three entry levels and two higher levels:

- **Level 2 (L2)**, equivalent to a school age pupil who could attain GCSE grade A-C.
- **Level 1 (L1)**, equivalent to a school age pupil who could attain GCSE grade D-G.
- **Entry Level 3 (E3)**, equivalent to school age pupils with special needs whose attainments fall below GCSE grade G.
- **Entry Level 2 (E2)**, between E1 and E3.
- **Entry Level 1 (E1)**, beginner level.

The ALCC contains guidelines on what should be taught at each level. Furthermore at each level, guidelines are specified for three areas: Text Focus, Sentence Focus and Word Focus. Text Focus correlates roughly with the discourse level in linguistics. Text Focus teaching guidelines include teaching an adult beginner: what a text is; the different purposes of texts; how different types of texts are organised; how to find information in different text types; how to use background knowledge to predict content and infer the meaning of unknown words; and how to use knowledge of spoken language to help with reading.

ALCC Text Focus guidelines include discourse connectives. Simple temporal discourse connectives (e.g. 'then') are introduced at E2 and causal connectives (e.g. 'so') at E3. This information suggests that E1 learners might not cope with connectives at all, whilst L1 and L2 learners may be able to cope with increasingly complex connectives. The ALCC is a rich source from which to build knowledge into our NLG system along with other sources such as existing psycholinguistic evidence and data from our own experiments. We can also use the ALCC guidelines to build hypotheses for our experiments.

Related work

Others have attempted to 'simplify' documents for various purposes in the past. Starting with an existing document, that may, or may not, be 'hard' to begin with, the document is modified to 'simplify' it. Past attempts to 'simplify' in this way have used a variety of methods. Examples are substituting common words for uncommon words (Devlin and Tait, 1998), activising passive sentences and resolving references (Canning, 2000) and reducing multiple-clause sentences to single-clause sentences (Chandrasekar and Srinivas, 1997), (Canning, 2000). Devlin and Canning's system, PSET, was aimed at simplifying documents for aphasic readers.

In addition, several past NLG systems tailor documents according to assumed background knowledge of the reader, for instance POWER (Dale et al., 1998) selects common words and phrases for naïve readers and technical ones for experts.

These systems use a two-level model of simplification, that is: 'naïve' vs. 'expert', or 'simple' vs. 'complex'. But they do not formalise what the two levels of the model mean. They do not explain what 'simple' and 'complex' mean. The PSET system for aphasics is the most comprehensive we know of, since it did use existing psycholinguistic evidence on which particular linguistic features aphasics find difficult. It was also evaluated with aphasic readers. However, the aphasic readers were not subdivided into different ability groups and there are many different types of aphasics.

As far as we know, there has been no research to date in NLG aimed at tailoring documents for different reading levels, or for tailoring for more than two levels. This project aims to generate documents at many levels, not just two. We are evaluating the system with people whose reading ability has been measured (see below) so we have a great deal more information about what each reader can cope with than the PSET model did. Also we will use the ALCC, existing psycholinguistic data and data from our own experiments to build a model of the reading levels we are aiming to generate texts for.

Another related concept is that of readability. Readability formulae such as Flesch (Flesch, 1949) and Flesch-Kincaid were developed as tools to measure reading ease. They are calculated from measures such as word length, sentence length, etc. The theory being that the shorter these lengths, the more 'readable' the text. They can be used as predictors of Reading Grade, but only if they are provided with a well-written document as input (Powell, 1981), (Redish, 1981). So they can be useful only for expert human writers who are very skilful in modifying their documents to fit the length criteria. Readability formulae cannot give us information on how to generate documents at lower or higher reading grades. The only use of readablity formulae within this project would be as part of an evaluation toolkit after a document had been generated. However, other evaluation methods, such as trials with real readers (see below) are far more informative and are the only ones of any real value.

Other work includes that of Brown (Brown, 1994). Brown claims that "Any genre can be made cognitively simple or more difficult by the manipulation of a small range of parameters..."

(p.14). According to Brown, the factors affecting cognitive load depend on: "number and distinguishability of referents"; "simplicity or complexity of spatial relationships"; "simplicity or complexity of temporal relations"; "simplicitiy or complexity of intentional/causal relations"; and "identity, similarity, or incompatibility of new information with that already available to the reader". A number of these are concerned with discourse relations and discourse connectives and are relevant to this project. The problem is that we need more than just a list of the parameters involved. We need to know how these parameters interact with one another and what the effects of modifying one parameter will be on all the others.

Developing preliminary hypotheses

We began our project by attempting to develop an initial hypothesis that we could test in a pilot experiment with real readers. Here we describe how we developed our first hypothesis. Consider these examples:

a. **Although** spelling is difficult, it is important.

b. Spelling is difficult, **but** it is important.

c. Spelling is difficult. It is important.

(a) can be 'simplified' by removing the discourse connective 'although' from the beginning of the first clause and inserting the more common 'but' at the beginning of the second (b). Data from a corpus of texts simplified by experts showed they made exactly these changes. Both (a) and (b) contain similar relations according to Knott and Dale's hierarchy (Knott and Dale, 1995), and they appear roughly equivalent in meaning. However, Oberlander and Moore (Oberlander and Moore, 1999) found that reading speed and comprehension (for good readers) are affected by changes in selection and position of connectives. Discourse analysts normally assign greater importance to certain text spans in a document over others. Changing position and selection of connectives in this way could possibly upset readers' perception of relative importance. Both of these should be taken into consideration and we should be wary that substitution of a common word for a less common one may not actually increase readability.

'Simplification' (c) removes the connective completely and splits one sentence into two. The result is less cohesive and the rhetorical relation must be inferred. There is psychological evidence to support the theory that good readers remember better when they have to infer relations. Extra mental work is required to build up a mental model of the meaning. This extra effort aids memory retention (Millis et al., 1993). If this is so, then perhaps discourse connectives make reading too easy for good readers and they would perform better if they had to infer. However, there is also a danger that they could infer the wrong meaning. On the other hand, Degand et al (Degand et al., 1999) observed that removal of even a few connectives from a text affects comprehension and memory of the entire content. Also, inferencing might place too great a mental load on bad readers, since they already have to expend greater effort on decoding words. If this is so, then connectives will help bad readers more than good readers.

In our pilot experiment, we decided to concentrate first on the presence and absence of discourse connectives. We tested an initial hypothesis that presence of connectives will help bad readers more than good readers. In future experiments we can go on to test other factors such as connective selection and positioning.

System design through experimentation

In this project, as well as utilising existing theories in our system design, we are also attempting to inform the design by testing it with users. System evaluation with real users often occurs only at the end of development when it is too late in the project to make changes to the design. In this project, we intend testing and evaluation to be an ongoing part of development. We are focussed right from the beginning on the end-users of the system, that is, the readers of the texts the system generates. It is important to trial with people who will actually read these documents. Since our system generates feedback reports for adults who are learning to read, these are the people who we asked to trial the system. We cannot know, until we do this, how difficult these readers will find the texts. Our pilot trial, described below, was designed

as a psycholinguistic experiment with adult basic English learners as participants. They are representative of the seven million people in the UK population who have difficulty with reading and writing. The only difference is that these people have recognised that they need to improve and have enrolled on a course in order to achieve this. It turned out that observations and discussions with these people and their tutors proved, in the case of this pilot, to be more informative than the actual experimental data obtained.

Implementation

The NLG System

We took an existing NLG system (Reiter et al., 2000) that generates feedback reports following a literacy assessment. The system's inputs are sets of answers from the literacy assessment in the Target Skills application by the Basic Skills Agency et al. (2001). An assessment contains up to 90 questions and each report is tailored specifically to the answers an individual gives.

Fig 1 shows an example of a document output by the system. Discourse connectives have been generated in the section above the horizontal line, but not in the section below the line. Note that the text above the horizontal line is the feedback report and is tailored for the individual who was assessed. The text below the line is fixed text.

The feedback describes an individual's strengths and weaknesses (see the sections headed 'Your results'). This section also contains a results chart that is a replica of the one produced in Target Skills at the end of the literacy assessment. The next section suggests how to improve reading and writing skills (see the sections headed 'What to do next'). The activities suggested are linked to the Target Skills application that has teaching materials at each level of the ALCC. The NLG system was originally designed to be part of Target Skills' pilot system, Skillsbuild. Typical content and style for the feedback sections of the report were obtained from knowledge acquisition carried out with adult literacy tutors. The final section of the report contains fixed text with general information about Target Skills (see the section headed 'Target Skills'). This section was included specifically for the purpose of comparison in reading speed and comprehension experiments.

We modified the NLG system to generate paragraphs either with, or without, discourse connectives. For each paragraph, a flag is set if discourse connectives are required. A small set of discourse connective insertion rules are employed, and additional clause-conjoining rules, where necessary.

The core of the feedback report generator is based on Reiter and Dale's architecture (Reiter and Dale, 2000). Each document has a fixed structure based on the four sections described above. The content for each section (except the last, which contains fixed text) is selected according to a set of rules about answers an individual gives during the literacy test. The rules were derived from knowledge acquired from adult literacy tutors.

Our knowledge acquisition process has been ongoing from the time the original NLG system was developed in 2000. During 2000, five Adult Literacy tutors and experts collaborated during system development and a further six tutors during the trials. During 2001, five more tutors have collaborated. Knowledge acquisition has involved recorded think-aloud sessions, sample feedback report writing, face-to-face and e-mail discussions, written questionnaires and web-based questionnaires. A total of sixteen experts have been involved throughout the UK in Aberdeen, Bourneville, Carshalton, Cambridge, Christchurch, London, Newcastle, Salisbury, Southampton and Telford.

Adult literacy tutors had some difficulty with sample report writing because feedback is normally given verbally to adult learners, not in writing. They gave suggestions about content and gave examples of language style and phrases that were incorporated into the system. The content of reports currently generated consists of informative factual information about scores, explanations of good and bad performance, and suggestions for activities to improve skills.

Tutors agreed that the purpose of feedback should not only be to inform on a learner's performance, but also to build confidence and encourage the learner to improve his/her literacy skills. However, it is not easy to evaluate whether a report will achieves goals such as 'encourage' or 'build confidence'. Sometimes at-

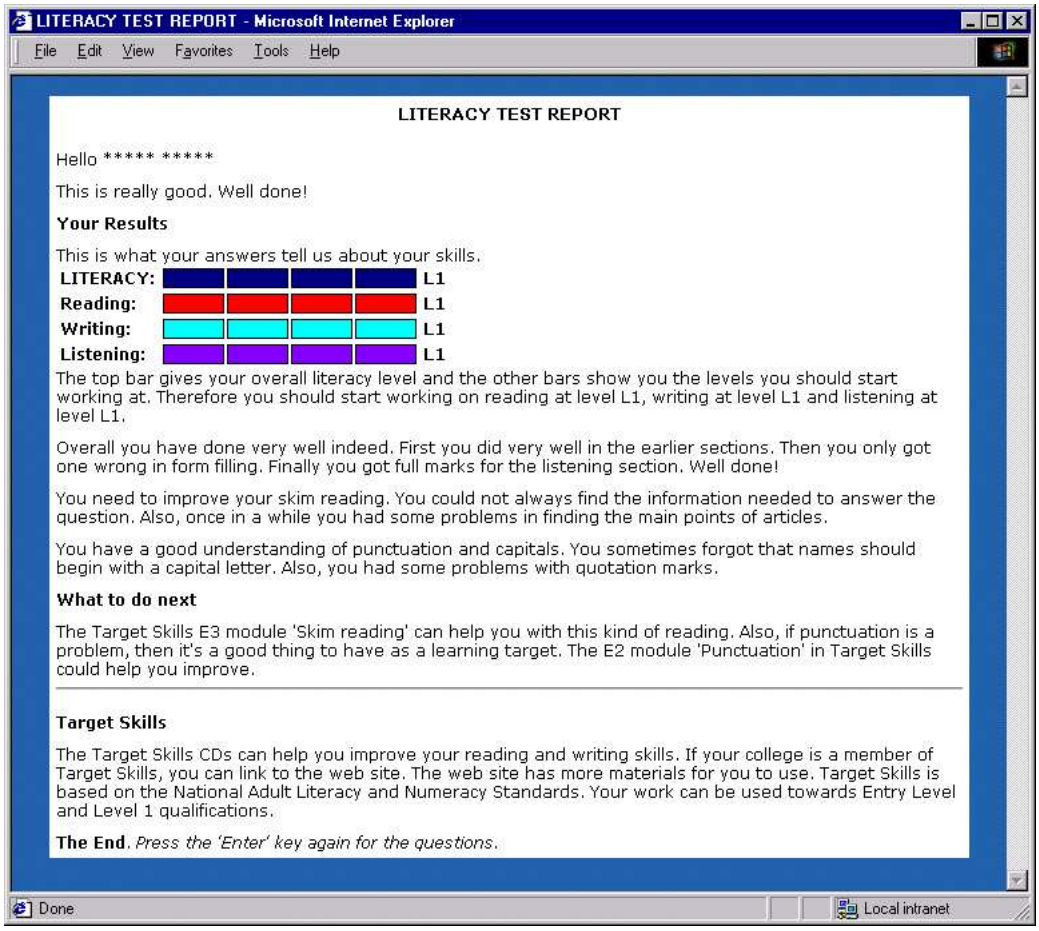

Figure 1: A literacy assessment feedback report generated by the NLG system

tempts to do this backfire. In the STOP project (Reiter et al., 2000), some people found an 'empathetic' style of feedback patronising. The perception of a document as patronising probably varies from person to person. There may also be a large overlap between encouraging and patronising styles and different people might draw the line between the two in different places.

Some people might even find 'depersonalised' reports more encouraging. An example would be feedback that includes appropriate quotes from other learners who had been in a similar position, but who have achieved success. This would demonstrate that other people have similar problems or experiences and it might avoid the 'patronising text' problem

Adult literacy tutors sometimes gave information that was contradictory. For instance, when asked if personal learner differences such as age, gender, and personality should be taken into account when generating feedback, tutors disagreed. They were asked: Suppose two people give an identical set of answers to a test, should their feedback differ? One said feedback should be varied with a learner's level of confidence or motivation. Another said it would be highly unprofessional to vary feedback according to personality traits. Another's opinion was somewhere between the former two in that he would give the same basic factual feedback to every learner, but if he were to give the feedback face-to-face, then the illustrative examples he would use might vary from person to person.

What do readers of literacy assessment feedback reports require? Do they want texts that they can read and understand fully? If so, texts for very low reading levels might have to leave out important information that would be in-

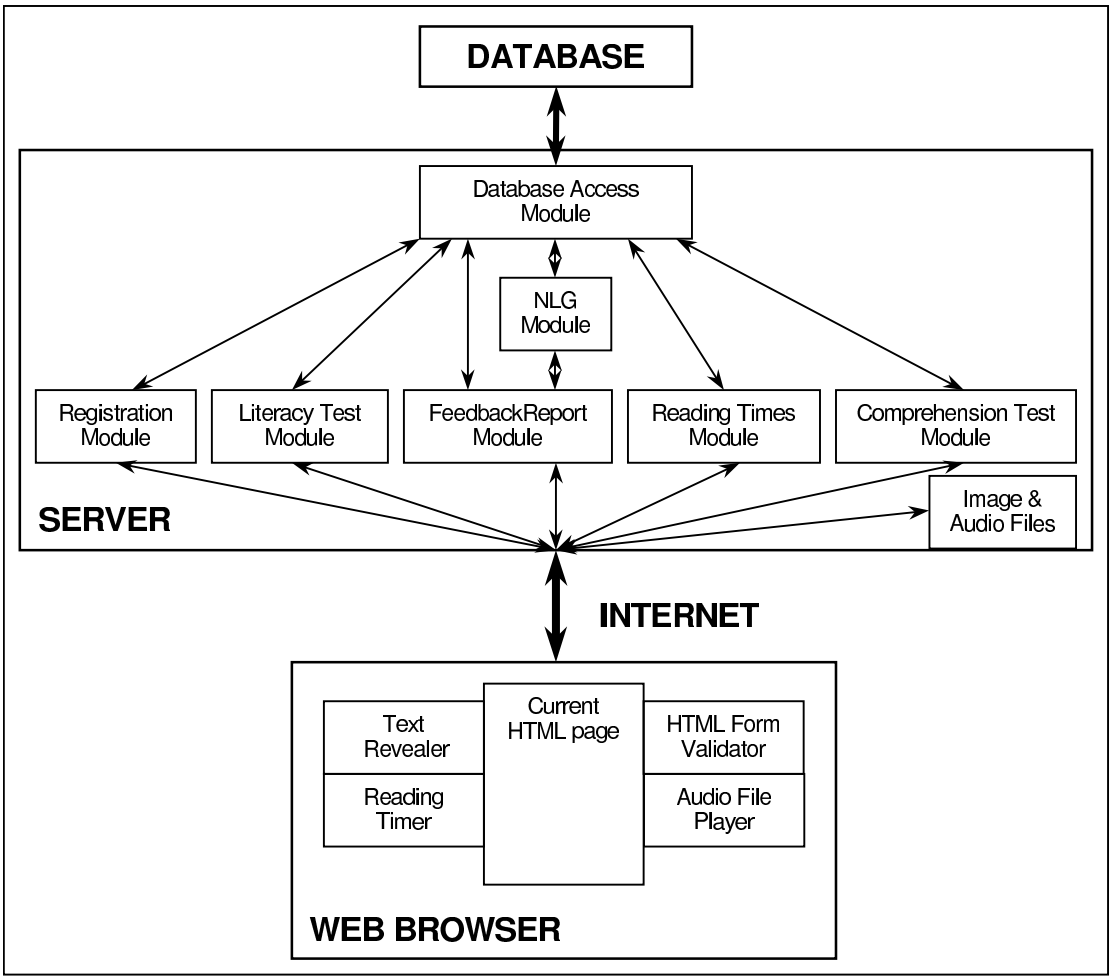

Figure 2: System architecture diagram

cluded for higher reading levels. For instance, a feedback report could leave out a detailed discussion of errors made and some suggestions for how to improve skills. On the other hand, learners might appreciate all the feedback the system can give, even if it takes them a great deal of time and effort to read it. An early user trial of the literacy assessment report generator in 2000 showed that adult learners wanted as much information as possible about their results. In fact some wanted details about their answers to every question.

Our knowledge acquisition process remains ongoing. As we trial each new version of the NLG system, we get an opportunity to discuss the content and language of the feedback reports with adult learners and their tutors. Tutors discuss the experience with their class after trials and give us further ideas and suggestions. Both of these occurred during the pilot experiment described below.

Test-bed Application

The NLG system is part of a larger testing application developed for this project. It includes software for participant registration, literacy testing, report generation, reading time testing, comprehension testing, and database access. The application is web-based. The decision to use a web-based architecture was made because we needed the system to be accessible from a variety of different places and machines as we travelled to Further Education Colleges and Adult Education Centres to trial the system. Colleges place severe restrictions on installing new software, and with this arrange-

ment, we could access our software via common web browsers without installing anything. Also all the experimental results could be conveniently collected and stored in a single relational database.

The system architecture diagram in Fig 2 shows the following server modules:

- **Registration Module.** This processes an HTML registration form containing information about a participant's name, id, age, gender, college, eyesight, hearing, education level, first language and anything known to affect reading (e.g. if the participant is dyslexic). These details are sent to the database access module which stores them in the database.
- **Literacy Test Module.** This administers the Target Skills literacy test mentioned previously. The literacy test was re-implemented so that each of the ninety questions is generated one at a time as a form on an HTML page. This module also sends the answers received from each participant to the database access module which stores them in the database.
- **Feedback Report Module.** This instructs the NLG system to generate a report. It randomly chooses to include discourse connectives or not. This choice is recorded in the database.
- **Reading Times Module.** This processes a cumulative string of reading times returned by the client-side reading timer. It retrieves each time from the string and sends the times to the database access module to be stored in the database.
- **Comprehension Test Module.** This administers the comprehension test in a similar manner to that of the literacy test module.
- **Database Access Module.** This sends SQL queries to the relational database. A large amount of data for each participant can thus be stored conveniently in database tables. Data can be cross-referenced and accessed as required.

Important distinctions are made between processing that takes place on the server and processing that takes place on the client's web browser. There are a number of tasks that it is more efficient to process on the client's web browser. These are:

- **Text Revealer.** This presents sections of the report during timed reading.
- **Audio Player.** This displays audio player controls in the web page. An audio file plays automatically when question web pages are loaded and it gives instructions about what the participant should do. The participant can click on the controls to replay or stop.
- **Reading Timer.** This uses functions to time in milliseconds starting from when a new part of the text in a page has been revealed. Times are added to a cumulative string and the string is sent to the server only when reading has been completed.
- **HTML form validation.** This is performed in the standard way, checking strings that have been entered, or whether radio buttons have been clicked. Functions are provided to bypass the validation if the user wants to skip a question or exit temporarily or permanently from the experiment.

Using client-side processes cuts down on time spent downloading or uploading from/to the server.

Pilot Experiment

Participants in the experiment were eighteen adults aged 17–65 attending adult basic skills courses at Southampton City and Salisbury Further Education Colleges and Christchurch Adult Education Centre. Two participants were dyslexic, one had unspecified learning difficulties, and one had brain damage as a child. Others had no known impairments affecting reading. Two were second language learners. Native English speaking university students were also included as 'expert readers', for comparison.

First, reading levels were measured using a re-implementation of the Target Skills literacy assessment mentioned previously. This is a skills-based literacy test. It grades readers into the five categories used in the ALCC and described

earlier. Feedback reports were then generated. Half the participants received discourse connectives in the tailored part and none in the fixed part, the other half received the reverse. Reading times for each heading and paragraph were measured. Finally, comprehension questions were administered and the answers were recorded and timed.

The numbers of valid results obtained were too few for statistical analysis. Time taken for participants to complete the experiment was longer than expected. The learners were using time allocated during their normal weekly class. Some only had an hour and a quarter and this was not long enough. Others who completed had a full two hours. Of the eighteen adult basic skills learners who participated, only eight completed the literacy test. Two people closed a window by mistake, thus losing their reading times. Only six sets of valid reading time results have been recorded to date. Of these six, two had literacy assessment results ranked at Entry Levels and the other four at Level One. Reading times for three 'expert' university student readers were recorded for comparison.

The rather sparse results show reading times per syllable in the range 50–500 ms with generally longer times for the 'expert' readers and generally longer times for the sections containing connectives. Reading times per word reveal similar longer times. This is not what we expected.

The longer reading times for 'expert' readers leads us to suspect that some learner readers may not have been reading the text 'properly' before pressing the button for the next part. Perhaps they were skipping parts that they could not read. The poor overall results for the learners in the comprehension test confirms that parts of the report were not understood. In our next experiment, we are considering asking participants to read the feedback aloud and making an audio recording. Hesitations and pauses may then indicate if words and phrases cause difficulty. It is also possible that intonational analysis could reveal whether readers are making mistakes when fitting new information in with what has gone before.

Answers to the comprehension questions were mixed. The learners answered 54% of the questions correctly, the experts answered 71% correctly. Mean times to answer the comprehension questions (including times for re-reading the report) ranged from 3.5 seconds to 13.5 seconds per question. There were no obvious time differences between 'expert' readers and learners.

In our next version of the system, we plan to generate feedback after each skill test. This will mean that we will obtain some results for each participant even if he/she does not have time to complete the entire literacy test. Even if a participant only gets half way though the literacy test, we will have an indication of his/her literacy level.

We received some very useful comments from the adult learners and their tutors. It is clear that there is domain knowledge the learners simply do not possess. For instance, some words and phrases used in the reports were unknown to adult learners because:

- they are technical terms in basic skills teaching that learners could not be expected to know, e.g. 'L1', 'Level 1', 'Skimming', or
- they are domain terms for skills that learners had not come across yet in their reading and writing education.

Part of the problem was due to our replication of the Target Skills Results Chart which contains technical terms and this could be removed. The activities in the literacy test must be referred to and should do this without using technical or domain terms for some readers and without using long explanations such as 'the part where a window popped up for a short time and you had to read through it really quickly to find a piece of information'. If the next version of the NLG system generates feedback after each skill test, as proposed above, then it could simply refer to 'the questions you just did'. Even if these changes were made, our system would still requires a better model of what readers at different levels might be expected to know, and more importantly, what they might be expected **not** to know.

There was another problem due to imperfect modelling of readers' knowledge. One tutor did not like the way that a learner had punctuation described as a "weak area" in "Your only weak

area is ...". This phrasing was derived from a sample report written by a tutor who had collaborated with us earlier. The tutor who did not like it felt that it sounded as though it was the learner's fault. The learner had never been taught punctuation and had never heard of it, so it did not make sense to describe it as "weak". The tutor suggested that the report should be worded more carefully and not assume knowledge that the learner may not have.

Lastly we will mention an issue that is important for any discourse level researcher. Features at the discourse level are dependent on lower level features. It does not make sense to try to make measurements on high level discourse features if participants have problems with lower level lexical and phrasal features as ours did. Errors will occur. Most researchers experiment only on good readers and do not encounter this problem. Of course we can change the words and phrases used by our NLG system, but we cannot ensure that our readers will not encounter some that they do not understand.

Conclusions

In this paper we have attempted to make the following points about our project:

- Little is known about how to automatically generate documents for different reading levels. If this project can discover more, then it will help both readers and writers.
- This project draws from a wide range of sources including education, psychology, linguistics and computer science.
- It is important to build formal models of reading levels, rather than simply labelling them as 'easy' or 'hard'.
- It is clear that a complete model of readers would be able to reason about the domain knowledge readers are likely **not** to have, as well as domain knowledge they are likely to have.
- Development of our NLG system is through experiments and trials as well as existing theories.
- Knowledge acquisition from tutors and adult learners is an ongoing part of system development.
- We have built a testbed application for experiments and trials.
- It is important that experiments and trials should be carried out with real users.
- We have undertaken a pilot experiment that raised many useful issues and has led to plans for improvement. Two important experimental issues arose. Adult beginner readers may not actually 'read' a document when they say they have. They may skip parts they cannot cope with. Features at the discourse level depend on lower level features and participants may have problems with these that affect the results.

Acknowledgements

Grateful thanks to my supervisor, Ehud Reiter. Thanks also to Yaji Sripada, Liesl Osman, Colin Gray, Rosemary Stevenson, John Tait, Martin Good, Jo Dixon-Trifonov, Val McKenzie, Jill Attewell, Jim Binning, Marlene Prangnell, Gang Lei, Bruce Scharlau, Stuart Chalmers, Nick Murray, Michael Ritchie, and John Martin. This work was supported by the Engineering and Physical Sciences Research Council (EPSRC).

References

Gillian Brown. 1994. Modes of understanding. In *Language and Understanding, Kirsten Malmkjr, Alastair Pollit, John Williams (eds.)*. Oxford University Press.

The Basic Skills Agency BSA, CTAD, and ASE. 2001. *Target Skills: Initial Assessment CD, version 1.* The Basic Skills Agency, 1-19 New Oxford St, London.

Y. Canning. 2000. Cohesive simplification of newspaper text for aphasic readers. In *3rd CLUK Colloquium.*

R. Chandrasekar and B. Srinivas. 1997. Automatic induction of rules for text simplification. *Knowledge-Based Systems*, 10:183–190.

Robert Dale, Stephen J. Green, Maria Milosavljevic, Cécile Paris, Cornelia Verspoor, and Sandra Williams. 1998. Dynamic document delivery: Generating natural language texts on demand. In *DEXA Workshop*, pages 131–136.

Liesbeth Degand, Nathalie Lefèvre, and Yves Bestgen. 1999. The impact of connectives

and anaphoric expressions on expository discourse comprehension. *Document Design: Journal of Research and Problem Solving in Organizational Communication*, 1:39–51.

Siobhan Devlin and John Tait. 1998. *The Use of a Psycholinguistic Database in the Simplification of Text for Aphasic Readers.* CSLI Publications.

Rudolph Flesch. 1949. *The Art of Readable Writing.* Harper, USA.

Alistair Knott and Robert Dale. 1995. Using linguistic phenomena to motivate a set of coherence relations. *Discourse Processes*, 18(1):35–62.

K. Millis, A. Graesser, and K. Haberlandt. 1993. The impact of connectives on the memory for expository text. *Applied Cognitive Psychology*, 7:317–339.

C. Moser. 1999. Improving literacy and numeracy: A fresh start. In *Report of the working group chaired by Sir Claus Moser.*

Jon Oberlander and Johanna D. Moore. 1999. Cue phrases in discourse: further evidence for the core:contributor distinction. In *Levels of Representation in Discourse, Working Notes of the International Workshop on Text Representation, Edinburgh University*, pages 45–54.

Kenneth Powell. 1981. Readability guides are helpful if ... *IEEE Transactions on Professional Communication Forum: Readability Formulas: Used or Abused?*, 24.

Janice Redish. 1981. Understanding the limitations of readability formulas. *IEEE Transactions on Professional Communication Forum: Readability Formulas: Used or Abused?*, 24.

Ehud Reiter and Robert Dale. 2000. *Building Natural-Language Generation Systems.* Cambridge University Press.

E. Reiter, R. Robertson, and L. Osman. 2000. Knowledge acquisition for natural language generation. In *Proceedings of the First International Conference on Natural Language Generation (INLG2000)*, pages 217–224.

Andrew Steeds. 2001. Adult literacy core curriculum including spoken communication. volume ISBN 1-85990-127-1. CTAD Ltd and The Basic Skills Agency.