Part of an evaluation of a natural language generation (NLG) system with real users, this study investigated whether readability rules used by the system were effective or not. We found that adults with poor reading skills made fewer substitution errors (miscues) reading aloud system output that had been generated when the rules were activated. So the rules seem to be at least partially successful.

**Background**

- If a text is more readable, readers make fewer reading errors and fluencies.
  - Oral reading errors increase as text difficulty increases (Bleschall and Wills 1984; Young and Bowers 1995).
  - Poor readers’ ability to recognize phrase boundaries decreases as text difficulty increases, but good readers’ ability remains unaffected (Young and Bowers 1995).
- NLG system, GRLU (Sandra Williams’ PhD project).
  - System generates a report after a literacy assessment.
  - System plans output using knowledge about discourse relations.
  - System uses readability preference rules, e.g. prefer solutions with:
    - Discourse cue phrases — to aid comprehension (Cogley 1999).
    - Short, common cue phrases — more readable (Williams, Baez and Moll 1988).
    - Evolve sentence punctuation — shorter sentences improve comprehension (Schriver 1982).
    - End-of-phrase punctuation.
  - Order with a statement first then related information.
  - Using the rules, the system would select [a] from the example below — “easy” version.
  - The system has rules base on frequency to generate an alternative text — “hard” version.

**Hypotheses**

- Poor readers will make fewer reading errors when reading an “easy” version of a report generated by the system than a “hard” version.
- It will make little difference to good readers which version they read.

**Participants**

- 36 participants.
- 19 good readers, score = L2 in literacy test, 17 poor readers, score < L2.
- Aged 16 years to over 60 years.

**Method**

1. Administer Literacy Test

2. System Automatically Generates a Feedback Report (version chosen randomly)

3. Digitally Record Reading Aloud

**Classifying reading errors**

- Substitutions
- Insertions
- Before or mid-word, “slurges” (Hubladder et al 2002).
- Misread (insertion) followed by correct word, i.e. self-corrections (Hubladder et al 2002, van Hassel 2002).
- Pauses
- Pauses before- or mid-word, i.e. reading disfluencies (Hubladder et al 2002).
- Omissions
- Words or syllables missed out (van Hassel 2002).

**Results Summary**

- Poor readers — More substitution errors reading hard texts than easy texts.
  - Means of 0.4 errors on easy texts and 2.3 errors on hard texts.
  - Significant variations in distribution (p < 0.001; Lawer’s test).
  - But not enough data.
  - Insertion errors — no significant differences.
  - Pause errors (differences) were longer reading hard texts than easy texts, but no significant differences.
  - Total errors (substitutions plus insertions plus pauses): no significant differences.
- Good readers
  - Pause error times were longer on hard texts (mean 432ms) than on easy texts (mean 101ms), but no significant differences.
  - Substitution errors — no significant differences.
  - Insertion errors — no significant differences.
  - Total errors (substitutions plus insertions plus pauses): no significant differences.
- 167 errors found (21 substitutions, 49 insertions, 96 pauses, 1 omission).

**Conclusions**

- Overall, substitution errors (traditional miscues) turned out to give the best evidence that our hypotheses could be correct and that the system is indeed generating more readable texts.
- Poor readers made more substitution errors on hard texts, so the NLG system’s rules for generating readable texts are working to some extent.
- The text version that was read made little difference to good readers. However, they were slightly more fluent (made fewer pause errors) on easy texts, indicating that perhaps the readability rules may help them too.
- The results indicate that the Natural Language Generation system has gone some way towards generating texts that are easy to read for poor readers. But we feel that further work is necessary to improve performance.

**Future Work**

- We will continue to investigate how to communicate with people who have poor reading skills in a new project, SkillSum. In this project, the research focus is on:
  - How to generate language to motivate people to take up basic skills courses.
  - How to generate language that is more readable.
- A fair proportion of reading errors were due to clusters of consonants and vowels as Labov et al found (1998). We will use this idea for improving lexical choice rules.
  - Prefer words that are easy to “round out” and pronounce.
  - Prefer words that don’t contain consonant and vowel clusters.