# AI for Improving Children's Reading Comprehension

Rosella Gennari[1] and Pierpaolo Vittorini[2]

[1] Faculty of Computer Science, Free University of Bozen-Bolzano, Piazza Domenicani 3, 39100 Bolzano, Italy
[2] Department of Life, Health and Environmental Sciences, University of L'Aquila, Viale S. Salvatore, Edificio Delta 6, 67100 L'Aquila, Italy

**Keywords:** technology enhanced learning; constraints; natural language processing; artificial intelligence; society; children.

## 1 The Problem and Solution

Developing the capabilities of children to comprehend written texts is key to their development as young adults. Nowadays, more than 10% of 7–10 year old children are poor comprehenders: they have difficulties in comprehending texts, e.g., reasoning on the temporal flow of a story. TERENCE (10.2010-09.2013) was an FP7 European project that developed the first adaptive learning system with learning material for primary-school poor comprehenders, made of stories and quiz-like games for reasoning about stories. The material is immersed in a game world and delivered in an adaptive fashion according to children's learning needs, e.g., reading skills, so as to promote a personalised experience. In particular, the TERENCE educational games are clustered and disclosed into levels of progressive reasoning (and language) complexity: factual games, concerning characters or their actions; time games concerning sequential or non-sequential temporal relations between events; causality games concerning causes or effects of events.

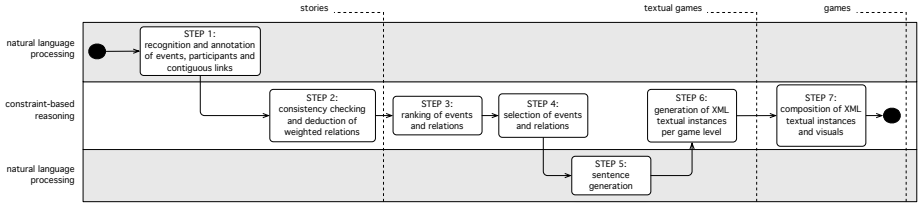
The TERENCE game world design adopted the evidence-based design for promoting a learning experience (i.e., improvement of comprehension) and the user centred design for promoting usability and, more generally, a positive experience. Now, the evidence-based design required c. 15 games of different complexity per story, for a total of c. 400 games per language. In order to tackle the issue in a general manner, TERENCE developed an AI-based process for generating games from stories, using natural language processing (NLP) and constraint-based technologies. The generated games were evaluated by education experts. Finally, the TERENCE system, its stories and games were used in a large-scale study at school for evaluating improvements in text comprehension. The paper outlines the semi-automated generation of games, the results of the evaluation of the generation process and the results of the large-scale study.

## 2 An AI-based Game Generation Process

Games contain a textual part with a quiz concerning a central event in the story and possible solutions to the quiz. All textual components come with visual representations. The game generation process interleaves three main software modules for generating the textual components and assembling them with the visuals: the story module; the textual game generation module; the visualisation module.

The story module uses NLP tools (for English and Italian) for recognising and annotating narrative semantics of texts: events, contiguous temporal and causal links between events, and entities that participate in each event. The story module uses constraint-based technologies for assigning a semantics to annotations, performing consistency checking and—in case of success—producing an enriched story through the deduction of non-contiguous relations between events, annotating additional game complexity parameters such as the distance in the text between events.

Taking in input the enriched stories, the game generation module performs several steps, again mixing constraint-based and NLP technologies. It starts by ranking events in the enriched stories according to the relevance of events for generating educational games of different complexity. It uses the ranking in order to select and then generate fixed numbers of game instances per game level, in XML. The visualisation module completes the work of the textual generation module by assembling the visuals and XML textual components. The workflow is depicted below.




## 3 Qualitative Evaluation and Impact

Once generated, games were experimentally evaluated using post-hoc human judgment. They were revised by education experts via a personalised interface and following a specific procedure for minimising human biases. Evaluators studied the text of a story and then revised all the games for it. For instance, they did a textual revision of grammar errors or for the selection of relevant events. They worked in pairs, their work was supervised by a moderator and tracked in structured diaries. A quantitative analysis of these showed a good quality of the game generation process, e.g., in only 5 cases it was necessary to change the event automatically selected through the ranking process as central for the game.

The major effort was concerned with the NLP-based generation of sentences. For instance, anaphora resolution in Italian was picked up as an additional effort

in the project for the game generation module. However, anaphora resolution in Italian is still a very challenging issue, e.g., 0-anaphora are frequent in Italian. Moreover, each quiz-game admitted only an item as solution, whereas the other items worked as distractors. Generating plausible distractors turned out to be another difficult issue. The documentation of the expert-based evaluation helped us in designing new heuristics for generating more plausible distractors, a future direction of work which seems promising for AI for quiz-like games.

Evaluating improvements in reading comprehension was the final and major goal of the TERENCE project. Improvements were assessed with a pre-post design large-scale study, with a control group. In the pre and post phases, all children's reading comprehension was assessed with standardised tests. The experimental group used TERENCE at school for circa 3 months, for about 1 hour a week; the control group followed normal school activities. According to our inferential statistical analyses, TERENCE significantly improves poor comprehenders' reading comprehension. That was the expected benefit that TERENCE promised for its learners. But the residual effects of TERENCE for the participants in the long-term evaluation are likely to go beyond text comprehension: organic functioning and global cognition (e.g., visual memory, attention), a potential increase in their communicative and relational competencies, as well as in their self-esteem.

### Additional information available on the web

TERENCE project web site: `http://www.terenceproject.eu/`; online demo: `http://hixwg.univaq.it/terence/3rd-release/`.