Cognitive and Metacognitive Educational Systems:
Papers from the AAAI Fall Symposium (FS-09-02)

Evaluations of the LODE Temporal Reasoning Tool with Hearing and Deaf Children

B. Arfé
University of Verona
Lungadige Porta Vittoria 17
37129 Verona, Italy
barbara.arfe@univr.it

R. Gennari*
Free University of Bozen-Bolzano
Via della Mostra 4
39100 Bolzano, Italy
gennari@inf.unibz.it

O. Mich*†
Free University of Bozen-Bolzano
Via della Mostra 4
39100 Bolzano, Italy
mich@fbk.eu

Abstract

LODE is a web tool for children that are novice readers, and is primarily meant for deaf children. It proposes written stories and interactive games for reasoning, globally, on the stories. In this paper, first, we motivate the rationale of LODE, and explain its reasoning games. Then we briefly describe the design of the web client-server architecture of LODE; the server employs a constraint programming system for creating and solving the LODE games in real time. Finally, we concentrate on two evaluations of the latest prototype of LODE: one with hearing novice readers; another one with deaf readers. We conclude by discussing the results of the evaluations, and their implications for LODE.

Introduction

Text comprehension enables a broad access to information and knowledge. In general, children develop adequate decoding skills in their first years of school. In shallow orthographies, such as Italian, text decoding is mastered early, by the age of 7 (Orsolini et al.). However functional literacy, that is, the ability to gather information and to learn from texts, implies more than decoding and parsing. It depends on the readers' detection and deduction of logical dependencies, and on the construction of a global mental representation of causal-temporal relations between the narrated events (Bamberg 1987; Trabasso and van den Broek 1985).

Children become sensitive to the role of logical connectives in narratives and to the global meaning of texts around the ages of 7 and 8 (Thompson and Myers 1985; van den Broek 1997). However, 7–8 olds are still novice readers as far as text comprehension is concerned. It is then, when children have still immature reading comprehension strategies and start developing them, that instructional intervention can be more effective.

The reading delay of deaf children is widely documented (Musselman 2000; Traxler 2000). According to the literature (Allen 1986; Furth 1966; Traxler 2000), their reading comprehension skills are usually those of novice readers, such as hearing 7–8 olds. Deaf children often tend to reason on isolated concepts and not correlate concepts in written texts (Arfé and Boscolo 2006). Such an attitude may also depend on the kind of "literacy interventions addressed to deaf children" that tend to "focus on single sentences and the grammatical aspects of text production" (Arfé and Boscolo 2006), whereas activities framed around storybook reading could support the deaf child's emergent and early literacy development (Schirmer and Williams 2003).

A novel literacy e-tool for novice readers, that is primarily meant for deaf children, should thus concentrate on neither grammatical aspects nor isolated sentences. Instead, it should stimulate children to reason, globally, on a story by correlating the narrated events. This is the main educational goal of our e-tool: LODE narrates written e-stories, and it also invites children to create simple e-stories. Through apt games, LODE asks children to reason on an entire story by deducing consistent relations among its events. A constraint programming system, embedded in LODE, is responsible for the consistency checking and the feedback to children.

In its current version LODE focuses on a specific type of deductive reasoning, namely, deductive temporal reasoning. Temporal dimension is a concept that children learn indirectly through narration. At the age of 5, normally developing children become able to make deductions with temporal relations, reasoning on sequences of events with "before" and "after" (McColgan and McCormack 2008). This ability seems to develop further from the age of 7 to that of 9 (Ge and Xuehong 2002), when children become able to master the "while" relation. The LODE's games use precisely the following temporal relations: "before", "after", "while". We refer to them as the LODE's relations henceforth.

This paper reports on and discusses two evaluations of LODE: an extensive evaluation with hearing 7–8 old novice readers; an experimental one with deaf children. The evaluations aim at assessing the feasibility of the LODE's reasoning games. They also compare two different graphical visualisations of an e-story's temporal relations. The work reported in this paper helps us frame and tackle two important issues: how to stimulate the development of global reading strategies in 7–8 old novice readers, and how to promote global and effective reading strategies in deaf readers.

*Work partially supported by the CARITRO Foundation, Trento, Italy, lode.fbk.eu.

†Also at FBK-irst, Povo (TN), Italy.

The remainder of this paper is structured as follows. It starts outlining the general architecture of the current prototype of LODE, specifying the role of the constraint system in LODE. Then it illustrates the reasoning games of LODE, and their rationale. Finally, it reports on and discusses the evaluations of the latest prototype of the tool.

The Architecture

LODE features a web-based client-server architecture. The server has a modular structure. The main modules are: 1) the e-stories' database, 2) the ECLiPSe constraint programming system (Apt and Wallace 2006). The stories' database of the current version of LODE contains temporally enriched versions of famous children's stories in XHTML format. Events and relations are temporally annotated in XHTML à la TimeML (timeML 2002); the main difference is that we do not restrict the relations between a pair of events to the atomic ones (Allen 1983). Such an extension of the language is necessary for correctly rendering inherently vague information.

The annotations of events and relations of a story are automatically modelled as a constraint problem in the language of ECLiPSe. The server architecture, in general, and the constraint module, in particular, are detailed in (Gennari and Mich 2007); in the following, we only provide a short update on them and the necessary information for this paper.

In essence, a constraint (satisfaction) problem is given by a finite sequence of variables, $x_1, \ldots, x_n$, each ranging on a domain D_i of values, and a finite set of constraints, namely, subsets of $\prod_{j=1\ldots m} D_{ij}$ with $0 \leq i_1 < \ldots < i_m \leq n$. A tuple of domain values $(a_1, \ldots, a_n)$ for the variables $x_1, \ldots, x_n$ is a solution to the constraint problem if it participates in all the constraints of the problem. Then we will say that a_i for x_i is consistent with the problem, $i = 1 \ldots n$.

In LODE, ECLiPSe solves a temporal constraint problem by deciding on the consistency of an Allen relation with the problem.

The client, a GUI, is an AJAX application compatible with most web browsers. It works as the interface between the LODE's user and the remote server. The general features of the GUI are described in (Gennari and Mich 2008a). The following section gives the essentials on the GUI and describes a type of games employed in the latest prototype of LODE.

The Comprehension Games

LODE narrates simplified versions of traditional children's stories, such as *The Ugly Duckling*, so that the language is more suitable to novice readers, and deaf children in particular. A story is divided into web pages as a storybook for young children. A demonstrator of the tool is made available online (lodedemo 2009). All the reasoning games of LODE use the LODE relations—see the introductory section. According to (Schaeken, V. der Henst, and Schroyens 2006), and in line with cognitive economy, people are often better with deductive tasks that admit solutions (a.k.a., models), and are happy with one plausible solution. The reasoning games of LODE are built on such findings, e.g., the games

Figure 1: LODE: a screen-shot of the CTx game with a positive feedback.

have only relevant textual information, and always admit a solution.

The latest version of LODE implements two main types of reasoning games requiring the constraint reasoner, namely, comprehension games and production games. In the remainder, we concentrate on the comprehension games, as these have been intensively evaluated with both hearing and deaf novice readers.

In the comprehension games, the child should choose one out of the three LODE relations connecting a pair of events of the story. Each event is described in words and by the related image in the story. The relations may be implicit in the story, thus children need to build a mental model of the story in order to choose the correct relation. Children will choose the correct relation if they have built a consistent mental model of the story temporal flow. The constraint system provides the feedback deciding on the consistency of the chosen relation.

LODE proposes two visualisations of the relations: one is more textual (CTx) (see Figure 1), the other is more graphical (CGr) (see Figure 2).

CTx In the CTx games, two images representing the events are described in words, and connected by three choice boxes. Each box proposes a written temporal relation (see Figure 1). We took such a visualisation with choice boxes from young children's books. Note that it is only the text that conveys the relevant information for answering the games, thus the child has to interpret the text in order to answer.

CGr In the CGr games, a temporal relation between two events is also rendered by the spatial position of the relative images along the timeline (see Figure 2). We propose such a spatial representation because several teachers for deaf children employ it. Note that it is not the text but this spatial linear representation of temporal relations that is conspicuous in the GUI, conveying information for tackling the games.

Figure 2: LODE: a screen-shot of the CGr game with a negative feedback.

The Design and Evaluation Main Phases

The design and evaluation of LODE is split into three main phases. We followed the user centred design (UCD) (Chadia, Maloney-Krichmar, and Pree 2004), and lately adopted methodologies of the participatory design (PD) in the evaluations with children (Markopoulos et al. 2008).

First phase The first phase is since long completed; it helped us to specify the context of use and the user requirements outlined in the introductory section. The users, considered in this phase, are two deaf children with their parents, and adults from different fields, given the multidisciplinarity of LODE. The adult users of the test were: two teachers for deaf children; two speech therapists; a linguist, expert of deaf studies; a cognitive psychologist, expert of deaf studies; five usability and accessibility experts. For more on the first phase, we refer the reader to (Gennari and Mich 2008b).

Second phase The second phase is over as well. We run two pilot tests concerning the usability of LODE with two groups of hearing children. Afterwards, we refined the LODE prototype, improving on its interface, simplifying the language and tasks of some of its reasoning games.

Third phase The third phase measures the usability of the latest LODE prototype, and in particular the feasibility of its reasoning games. The following section reports on two tests of the third phase:

(H test). the test with hearing 7–8 novice readers;

(D test). the test with deaf children.

The Evaluations of the Third Phase

Goals

In the evaluations of the third phase, we consider the following parameters:

- *time*: the seconds spent on the story, and the seconds spent on each game;
- *accuracy*: the number of trials/clicks per game before resolving it correctly.

The first parameter measures the time performance, and the second the accuracy in tackling the games.

The parameters serve our first goal, that is, to assess the feasibility of the comprehension games, developed with a constraint programming system.

They also allow us to carry on our second goal, that is, to compare the two visualisations of the temporal relations. Time performance indicates which visualisation can take children a longer cognitive processing. Accuracy can indicate which visualisation could induce a more effective comprehension processing.

Methodology

Participants Information concerning the children's literacy level and, in case of deaf children, their level of deafness were collected prior to the tests with the help of their educators.

H test. The *H test* with hearing children involved thirty-one children: seventeen girls and fifteen boys. Five children were 7 year old, the remaining children were all 8 year old. All the children were at the start of the third class of an Italian primary school. They had had the same teachers during their primary-school years, and hence they had received a similar literacy instruction at school. Only one child declared not to use the computer at home, whereas the other children said that they use it circa three times a week for about one hour. Twenty-four of them are not allowed to use Internet alone at home.

D test. The *D test* involved ten deaf children, aged 8 to 13, five boys and five girls. All participants are pre-lingually deaf. Seven are profoundly deaf from birth, two have a severe hearing loss, and one has a moderate hearing loss. Italian is their first language. All deaf children regularly use computer either at home or at school. However, seven of them do not regularly use Internet.

Procedure Both the H and D tests were run as field experiments (Markopoulos et al. 2008). During the tests, deaf and hearing children worked on Windows XP PCs, using Mozilla Firefox. Each child had a dedicated PC.

At the start of the tests, children were told that they only had to do what was written on their display and follow the interface arrows. The deaf child's preferred communication mode (either signed Italian or oral language) was used to present the tasks, and give instructions in the D tests.

Then each child got associated to a unique electronic ID and had to answer an e-questionnaire mainly concerned with their computer and Internet literacy.

At this point, each child could start the computer session, playing with LODE. A log file per child was created, collecting his or her time performance and accuracy as specified in the above section. The test participants had no time limit for completing the computer session; they were also allowed to stop before finishing the tasks. First, children had to read the *The Ugly Ducking* story, and then tackle the reasoning games. They could read the story more than once. The experimenter never intervened while children were playing with LODE.

Notice that children could see the forward arrow and move to the subsequent game only after correctly resolving the current one. They had six comprehension games to resolve:

1. CTxA, CTxB and CTxC, adopting the textual visualisation;
2. CGrA, CGrB and CGrC, employing the graphical visualisation.

The events used in the comprehension games were generated randomly from a set of seven extracted from the *The Ugly Ducking*. In this manner, time performance and accuracy are independent of the event types and orderings. Children were divided into two groups, TG and GT. The TG group tackled first the CTx game set, and then the CGr game set; whereas the GT group tackled the two game sets in inverse order. In this manner, we minimise the risk that the order of games can affect children's performances and accuracy.

H test. The TG group of hearing children was composed of fourteen children, while the GT group was composed of seventeen children. Originally, the TG group was composed of seventeen children as well, but three children had a severe language impairment, and inadequate text decoding skills.

D test. Both the TG and GT groups of deaf children counted five children each.

Results

Only five hearing children decided to read the story more than once. All deaf children read the story only once.

Resolution Time Analysis The *resolution time* of a game is defined as the time that elapses since the child opens the game page until the child clicks the forward arrow—remember that the child can see and click the forward arrow only after correctly resolving the game.

Let $t(Ex)$ denote the mean resolution time of the Ex comprehension game. The value of the mean time for CTx is $(t(\text{CTxA}) + t(\text{CTxB}) + t(\text{CTxC}))/3$, and the value of the mean time for CGr is $(t(\text{CGrA}) + t(\text{CGrB}) + t(\text{CGrC}))/3$.

In order to assess whether the difference between the resolution times of the CTx games and CGr games is significant, we conducted Analyses of Variance (ANOVA) with repeated measures. We considered the visualisation type as a within-subjects factor, since all the children had to resolve the same games, with two levels, textual or graphical.

H test. The average time of the computer session was about 25 minutes, out of which 6 minutes, averagely, were spent on the story. All children decided to read the story and tackled all the games. They spent averagely 24 seconds on Ctx games, and 43 seconds on CGr games. See Table 1.

	H test		***D test***	
	CTx	**CGr**	**CTx**	**CGr**
Mean	24	43	29	41
SD	12	24	25	27

Table 1: Resolution time in seconds.

H test		***D test***	
CTx	**CGr**	**CTx**	**CGr**
45%	61%	30%	40%

Table 2: Success rates

The main effect of the visualisation type was significant for the resolution time, with $F(1, 30) = 14.797$, $p < .0001$.

Thereby hearing children had a significantly longer resolution time with the graphical visualisation than with the textual one.

D test. The average time of the computer session was about 20 minutes for deaf children, out of which 5 minutes, averagely, were spent on the story. They spent averagely 29 seconds on Ctx games, and 41 second on CGr games. See Table 1.

The visualisation type is not statistically significant for the resolution time, with $F(1, 9) = 3.550$, $p = .092$. Deaf children took averagely the same time in resolving CTx and GGr games.

Success Rate Analysis As explained above, the comprehension games are 3-choice games. We consider a game correctly resolved if the child selects a consistent relation with:

- at most 2 choices in case of the CTxA and CGrA games, because these work as training games for the respective visualisation types;
- precisely one choice in the remaining games.

The success rate of a game set, CTx or CGr, is then equal to the percentage of children that correctly resolved all the games of the set. Table 2 displays the success rates of the CTx set and CGr set.

H test. A binomial test revealed that the success rate was above chance with $p < .05$ for the CGr game set, and only at chance level for the CTx game set.

D test. As Table 2 shows, deaf children tended to fail in both CTx and CGrx games. A binomial test revealed that success was approximately at chance level for this group, both in the CTx and in the CGr conditions.

Discussion

The average success rates of hearing children in the comprehension games gave us interesting information concerning the feasibility of the games. The success rate for the graphical comprehension games was above chance according to our analyses. Such games, resolved in real time with a constraint programming system, are thus feasible for hearing novice readers, 7–8 old.

Also the analyses on the resolution times and success rates of hearing children revealed some interesting results concerning the visualisation of the relations in the comprehension games. Hearing children took significantly longer in dealing with the graphical visualisation than with the textual one, see also Table 1. However, only the graphical visualisation gave success rates above chance, and better than the textual one, see also Table 2. Given this, we can suppose that the greater time spent on the comprehension games with the graphical visualisation induced a deeper elaboration by the child. The support of graphical information may have induced a deeper analysis of the text, improving their comprehension. This hypothesis is consistent with the assumption that pictures or graphical information are superior to words and textual information in semantic tasks (Dillon and Song 1997; Paivio 1991).

Deaf children failed in identifying temporal relations between two described events, both when they are represented graphically and textually. However, they tended to succeed more when such relations are expressed graphically, albeit not at a significant level. Resolution times did not differ between graphical and textual visualisation, indicating that deaf children spent equal time in elaborating the graphical and textual visualisations.

Summing up:

- the CGr condition seems to be more beneficial for hearing children. Resolution times show that the graphical visualisation of temporal relations (a linear visualisation) stimulate children to process the relations more, or more in depth than the mainly textual one;

- the exploratory test with deaf children does not show similar benefits of the CGr visualisation, whereas we expected them to benefit from it as well, or even more than hearing children.

Why so? A plausible hypothesis is that the two visualisations did not function differently for these deaf children. In particular, the CGr visualisation refers to a linear representation of time that we assumed to be rather intuitive and standard for them, but may be not. An alternative explanatory hypothesis is that reasoning with temporal relations on an entire e-story or temporal relations *per se* may be more difficult or require more training for these children. These children may take a longer time in building a mental model of the temporal relations between the illustrated events. Deaf children may also need more training with the LODE web tool in order to benefit from it. This is consistent with recent tests with a deaf child, who improved on the CGr games after using LODE repeatedly. It is also in line with the results of (Fajardo et al. 2008), which reports on different skills of deaf and hearing users in using web tools.

Conclusions

The paper outlined the architecture of LODE. For more details on this and the role of the constraint programming module, the interested reader is referred to (Gennari and Mich 2007). To the best of our knowledge, LODE is the first e-tool that employs an automated reasoner for reasoning, globally, on children's stories. For an overview of related literacy e-tools, see (Gennari and Mich 2008c).

A first design of the GUI and a preliminary evaluation plan were described in (Gennari and Mich 2008a). Then (Gennari and Mich 2008b) reported on preliminary pilot evaluations of the tool with two deaf children and with experts of usability or deaf studies.

The evaluations of the tool with hearing novice readers, presented in this paper, allowed us to gather relevant information concerning the usability of the tool and, in particular, the feasibility of the graphical comprehension games. More precisely, our analyses revealed that hearing novice readers took a significantly longer resolution time on the comprehension games with the graphical visualisation. However, their success rates with the graphical visualisation are above chance, and this visualisation seems to be more comprehensible for them than the textual one. Our results with deaf children suggests that they may need more training with LODE in order to benefit from the tool, its games and their graphical visualisations.

To conclude, this study sustains the feasibility of LODE and its temporal reasoning tasks for novice hearing readers. More research studies are needed with deaf users. Given the positive results of recent tests with a deaf child, a future evaluation session, with pre and post tests, will assess the learning efficacy of a mature prototype of LODE, in particular, its efficacy in promoting global reading strategies in novice readers, hearing and deaf.

References

Allen, J. 1983. Maintaining Knowledge about Temporal Intervals. *ACM Comm.* 832–843.

Allen, T. 1986. Patterns of Academic Achievement among Hearing Impaired Students: 1974 and 1983. In Schildroth, A. N., and M. A. Karchmer, e., eds., *Deaf children in America*. College Hill Press.

Apt, K., and Wallace, M. 2006. *Constraint Logic Programming using ECLiPSe*. Cambridge University Press.

Arfé, B., and Boscolo, P. 2006. Causal Coherence in Deaf and Hearing Students' Written Narratives. *Discourse Processes* 42(3):271–300.

Bamberg, M. 1987. *The Acquisition of Narratives. Learning to Use Language*. Berlin: de Gruyter.

Chadia, A.; Maloney-Krichmar, D.; and Pree, J. 2004. User Centered Design. In W. Bainbridge, e., ed., *Encyclopedia of Human-Computer Interaction*. Berkshire. 763–778.

Dillon, A., and Song, M. 1997. An Empirical Comparison of the Usability for Novice and Expert Searchers of a Textual and a Graphic interface to an Art-resource Database. *Journal of Digital Information* 1(1).

Fajardo, I.; Arfé, B.; Benedetti, P.; and Altoé, G. 2008. Hyperlink Format, Categorization Abilities and Memory Span as Contributors to Deaf Users Hypertext Access. *Journal of Deaf Studies and Deaf Education* 13(1):87–102.

Furth, G. 1966. A Comparison of Reading Test Norms of Deaf and Hearing Children. *American Annals of the Deaf* 111:461–462.

Ge, F., and Xuehong, T. 2002. Temporal Reasoning on Daily Events in Primary School Pupils. *Acta Psychological Sinica* 34:604–610.

Gennari, R., and Mich, O. 2007. Constraint-based Temporal Reasoning for E-learning with LODE. In *Proc. of the Thirteenth International Conference on Principles and Practice of Constraint Programming*.

Gennari, R., and Mich, O. 2008a. Designing and Assessing an Intelligent E-Tool for Deaf Children. In *Proc. of IUI 2008*. ACM Press.

Gennari, R., and Mich, O. 2008b. Evaluation of an Intelligent E-tool for Deaf Children. In *Proc. of ASSETS 2008*. ACM Press.

Gennari, R., and Mich, O. 2008c. Global Temporal Reasoning on Stories with LODE: a Logic-based E-tool for Deaf Children. Technical Report KRDB08-4, KRDB Research Centre, Faculty of Computer Science, Free University of Bozen-Bolzano, Italy.

2009. Lode, the demonstration. Retrieved June 11, 2009, from lodedemo.fbk.eu.

Markopoulos, P.; Read, J.; Mac-Farlane, S.; and Hoeysniemi, J. 2008. *Evaluating Children's Interactive Products. Principles and Practices for Interaction Designers*. Morgan Kaufmann.

McColgan, K., and McCormack, T. 2008. Searching and Planning: Young Children's Reasoning about Past and Future Event Sequences. *Child Development* 5:1477–1497.

Musselman, C. 2000. How Do Children Who Can't Hear Learn an Alphabetic Script? A Review of the Literature on Reading and Deafness. *Journal of Deaf Studies and Deaf Education* 5:9–31.

Orsolini, M.; Fanari, R.; Cerracchio, S.; and Famiglietti, L. Phonological and Lexical Reading in Italian Children with Dyslexia. Reading and Writing. In Press.

Paivio, A. 1991. Dual-coding Theory: Retrospect and Current Status. *Canadian Journal of Psychology* 45:255–287.

Schaeken, W.; V. der Henst, J.; and Schroyens, W. 2006. The Mental Model Theory of Relational Reasoning: Premises' Relevance, Conclusions' Phrasing and Cognitive Economy. In Schaeken, W.; Vandierendonck, A.; Schroyens, W.; and Ydewalle, G. E., eds., *The Mental Model's Theory of Reasoning*. Psychology Press.

Schirmer, B., and Williams, C. 2003. Approaches to Teaching Reading. In Marschark, M., and P. Spencer, e., eds., *Handbook of Deaf Studies, Language and Education*. Oxford University Press.

Thompson, J., and Myers, N. 1985. Inferences and Recall at Ages Four and Seven. *Child Development* 56:1134–1144.

2002. TimeML. Markup Language for Temporal and Event Expressions. Retrieved June 11, 2009, from www.timeml.org.

Trabasso, T., and van den Broek, P. 1985. Causal Thinking and the Representation of Narrative Events. *Journal of Memory and Language* 24:612–630.

Traxler, C. 2000. The Stanford Achievement Test, 9th Edition: National Norming and Performance Standards for Deaf and Hard-of-hearing Students. *Journal of Deaf Studies and Deaf Education* 5:337–348.

van den Broek, P. 1997. Discovering the Cement of the Universe: The Development of Event Comprehension from Childhood to Adulthood. In van den Broek, P., and Bourg, T. E., eds., *Developmental Spans in Event Comprehension and Representation*. Mahwah, NJ: Lawrence Erlbaum Associates, Inc. 321–342.