The manual revision of the TERENCE Italian smart games

Cofini Vincenza[1], Rosella Gennari[2], and Pierpaolo Vittorini[1]

Abstract TERENCE is an FP7 ICT European project, highly multidisciplinary, that is developing an adaptive learning system for supporting poor comprehenders and their educators. The paper describes the automatic smart games generation process in TERENCE, motivates the need for a manual revision and describes it in detail. The paper thus provides a thorough insight in understanding the quality level of the automatic smart games generation process in TERENCE, and the time/effort needed for their manual revisions.

Key words: Smart games, manual revision, NLP, reasoning

1 Introduction

TERENCE [10] is an FP7 ICT European project, highly multidisciplinary, that is developing an adaptive learning system for supporting poor comprehenders and their educators. Its learning material are stories and games. The games are specialised into relaxing games, which stimulate visual perception and not story comprehension, and smart games, which stimulate inference-making for story comprehension [8, 1, 2].

In brief, the TERENCE plan for stimulating story comprehension consists of the following increasingly demanding tasks [4]: firstly, it makes the learner reason about

[1] Cofini V, Vittorini P
Dep. of Life, Health and Environmental Sciences, University of L'Aquila, 67100, L'Aquila, Italy
e-mail: vincenza.cofini@cc.univaq.it
e-mail: pierpaolo.vittorini@univaq.it

[2] Gennari R
Faculty of Computer Science, Free University of Bozen-Bolzano (FUB), 39100 Bolzano, Italy
e-mail: gennari@inf.unibz.it

the characters that are in the story, then about the events that take place in the story, hence about temporal relations among the events, and finally about causal-temporal relations among events (see taxonomy in Fig. 1). Accordingly, factual, temporal, and causal smart games are the actual implementation of the corresponding comprehension tasks.

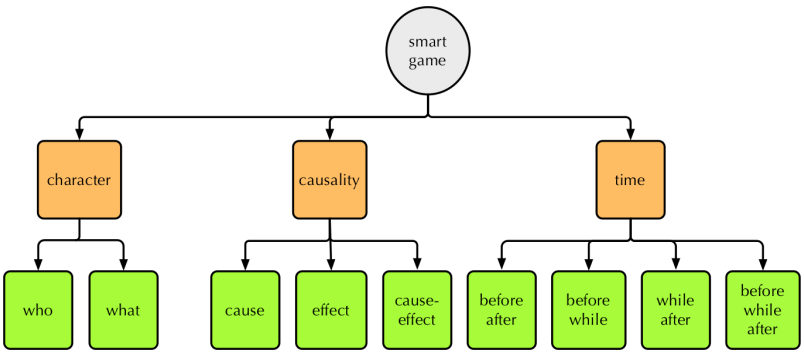

Fig. 1 Smart games taxonomy

In TERENCE, smart games are automatically generated as follows (see Fig. 2):

Phase A. Firstly, from a story text contained in the story repository, an NLP module generates a story annotated with a variant of the TimeML language, that was extended in [9] with tags that are relevant for the TERENCE smart games. For instance, the ENTITY and CLINK tags aim, respectively, at representing the entity related to an event, and the causal-temporal relations between two events. The annotated story is then stored in the same repository.

Phase B. Then, a reasoner checks the consistency of the annotations, detects the eventual temporal inconsistencies, and enriches the annotations by adding deduced temporal relations as further TLINK tags [6]. This new consistent and enriched story is also stored in the story repository.

Phase C. Starting from the consistent and enriched story, the reasoner module generates automatically instances of smart games. For instance, to create a WHO-game related to a certain event [7, 5]:

- the ENTITY that participates in the event with a role of protagonist is selected as the correct answer;
- other two entities that are not related to the event and are different from the entity selected above are added as wrong answers;
- the question asked to the learner is generated through a text-generation module (e.g. if the event is that "Ernesta is riding[1] a bike", the question will be "Who is riding a bike?").

[1] The verb "to ride" is detected as an event.

The resulting games are then stored into the game repository.

Phase D. Finally, a manual revision of the generated smart game instances takes place, where the related visuals (e.g. background illustrations, buttons) are also specified.

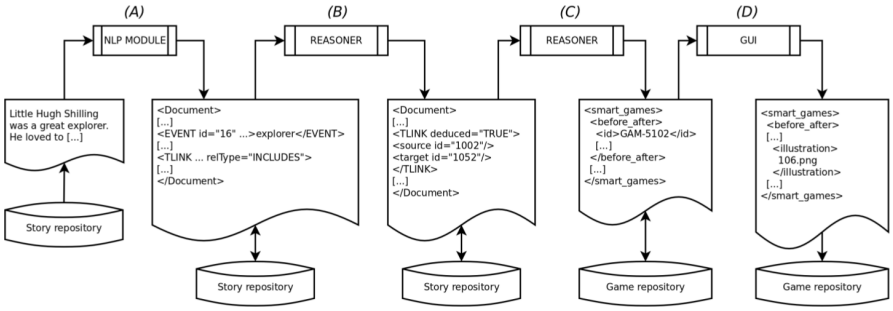

Fig. 2 The smart games generation workflow

It is worth remarking that the errors introduced in the automatic annotation process ("Phase A" mentioned above) influence the quality of all subsequent phases. For instance:

- badly recognised TLINKs (e.g. a BEFORE relation detected as an AFTER relation) lead to the deduction of wrong additional relations. Consequently, the smart games that includes such wrong relations may have temporal games with wrong solutions marked as correct;
- poor annotations may:
  - prevent the generation of some classes of games. For instance, without TLINKs with relation type of INCLUDES/IS_INCLUDED/OVERLAPS, the procedure will not be able to generate any BEFORE/WHILE, WHILE/AFTER, BEFORE/WHILE/AFTER smart game;
  - not offer enough alternatives for selecting plausible wrong choices. For instance, if only one entity is detected in the story, the wrong choices (see discussion on "Phase C" above) are taken from *other* stories with different characters. However, since these choices are not very much plausible, the quality of the resulting smart game is reduced.

As a consequence, the manual revision phase takes a crucial importance in the whole generation process. Understanding the amount of manual effort is then of a major interest (Sec. 2), since it indicates the overall effectiveness of the automatic approach used in TERENCE and may give better insight in the tasks that has to be carried out as priorities (Sec. 3).

2 Manual revision

The objective of the manual revision was to control the automatic generation and ensure the formal, technical, and content correctness of all the components of the game proposed for the stimulation of understanding of the written text. Details on the guidelines and on the software system that supported the manual revision can be found in [3].

The revision work was divided into 3 steps:

- Formal revision, i.e.: correction of grammatical and syntactic errors in the text, correction of punctuation, correction of the verb (present tense, active form), correction of referring expressions, check of sentence length and structure;
- Substantial technical revision, i.e.: check of game identification number, correction of the questions texts, correction of the solutions (by keeping fixed the main event), selection of new fixed events for solutions;
- Construction of cause/effect games, i.e.: text proposal, check out of proposals, games uploading.

Each operator studied the text of the story and reviewed all the games associated with it.

By proceeding on the basis of the output of the automatic generation, the manual review of TERENCE games was initiated through a software to speed up the audit work and its monitoring. The operator, accessing the software, could select the games by selecting the proper book and story. The Italian revision was conducted on 4 books: 3 books (16 stories) for students 7-9 years old and 1 book (9 stories) for students 9-11 years old.

The software shows on left of the interface the story text (so to let the operator to have the context always available) and enables in a tabular fashion to view the related games. After a game with its own ID was selected, it was possible to identify the type of game on the screen (before/after, before/while, before/while/after, what, while/after, who). Each game had its own *fixed event* (the event around which the entire game is build), completed with the event description, the game question, as well as a set of multiple-choice questions that depends on the selected game type.

The operator's job was firstly to verify the congruence between the fixed event text and the text of the proposed questions and solutions.

In general, the revision was structured, with the general principle to modify as little as possible the work of the automatic generation.

The revision consisted in correcting: (i) the grammatical and syntactic errors in the text; (ii) punctuation; (iii) verbs (present tense, active form); (iv) referring expressions; (v) sentence length and structure; (vi) the questions' texts [2]; (vii) the possible solutions; (viii) eventually the fixed event.

[2] With the aim of unambiguously identifying the event in the story. For instance, the same event text could be referred to different story episodes.

Each operator studied the story text, reviewed all the games associated with it and created from scratch the textual part of the causal games[3].

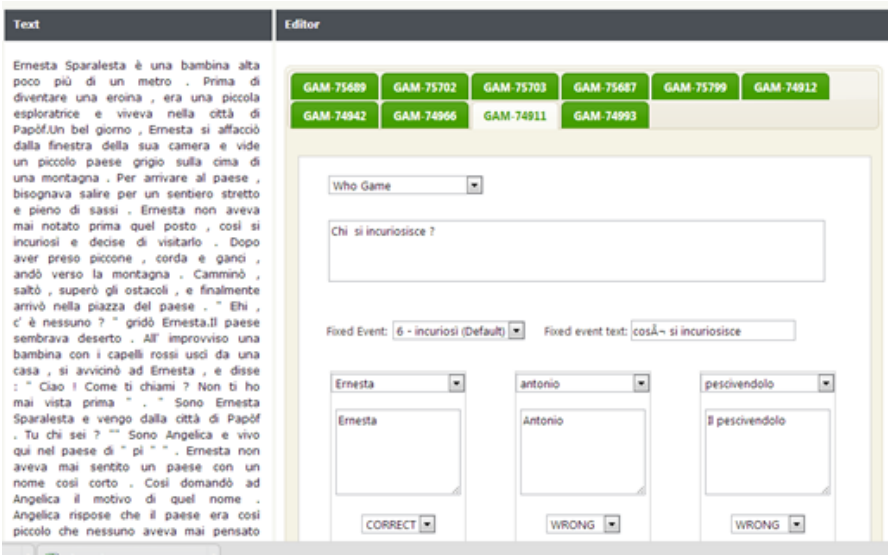

Fig. 3 Example of a review of WHO game (Book 2, Story 1)

In the revision of WHO-games, it sometimes happened that the solutions proposed were not consistent with the automatic generated question[4]. In such cases, it was necessary to change the solutions by choosing new entities from the drop-down menu and to guarantee the coincidence between entities and text. In the example shown in Fig. 3, the question asked is "Who is curious?". To make the necessary corrections, we had to (i) choose a new entity for each solution (by using the respective pull-down menu), and (ii) verify that it was the corret/wrong event actor.

We also had to take into account that each answer should be properly understood by all learners (poor comprehenders, deaf). Therefore, the changes were done by continually trying to work on subjects, preferring personal names, paying attention to the spatial distribution in the text and the kind of characters, and to not facilitate the reader in selecting the correct solution. Sometimes it was also necessary to completely change the game question associated with the fixed event, because the event was present several times in the text and associated with different subjects. In this case, it was also necessary to rewrite the question so to make it unambiguous.

In general, the review of temporal games was an even more challenging task, because it was necessary to locate the temporal coherence of proposed solutions on the basis of a fixed event, so that the wrong type solutions would exclude events

[3] The NLP module was in fact unable to detect any causal relation in the text.

[4] Probably due to errors introduced by the NLP anaphora resolution subsystem.

with overlapping time intervals in to the right answers. For instance, the following corrections were applied to a before/after game of Book 2, Story 1 (Tab. 1).

Solution	Pre-revision	Post-revision
AFTER	to thank	The inhabitants of the land of "pí" thank Jasmine
BEFORE	Louis leads the electrician the wires	No change
WRONG	All manage to split the fairly rubbish without difficult calculations needs	All manage to divide garbage in the right way

Table 1 Example pre/post revision for GAM 75702, Story 1, Book 2

Each operator had the task of filling in a diary (in a spreadsheet format, see Tab. 2), made up of 33 fields, all changes made in every revised game and depending on questions and solution types. This diary allowed the monitoring of all activities and their analysis as reported below.

Items	Choices
Operator name	
Book number	
Story number	
Game ID	
Kind of game	Who, what, before/after, before/while, while/after, before/while/after
Did you have to change the main question of the game?	Yes/no
If yes, write the new main question text	
Did you have to change the fixed event?	Yes/no
If yes, write the new fixed event text	
If yes, write the old fixed event text	
Did you have to change the automatic solution?	Yes/no
Kind of solution	Wrong, correct, before, after, while, cause, effect
Did you have to change the event associated with the automatic solution?	Yes/no
If yes, write the new fixed event as in the menu	
If yes, write the new text as in the box solution	
Notes	

Table 2 Revision games log

A total of 250 games were reviewed, with respect to 25 stories, with the highest proportion of games of type before/after (30%). Based on data reported by operators, the average review times were estimated. They were lower for reviewing a WHO-game, and higher for temporal games. On average, for a set of games, i.e. one game for each typology, it was necessary to work for circa 76 minutes. By considering the need for reading the related story, filling the excel, the average time of each operator

to finalise a set of games was equal to circa 90 minutes, i.e. approx 15 minutes per game (see Tab. 3).

GAME	n	%	Average time
WHO	25	10.00	10,6
WHAT	34	13.60	12
BEFORE/AFTER	74	29.60	12,8
BEFORE/WHILE	41	16.40	12,8
WHILE/AFTER	42	16.80	12,8
BEFORE/WHILE/AFTER	34	13.60	14,8
Total	250	100.00	75,8

Table 3 Details about the revised games

Only in the 6% of all cases, it was necessary to change the automatically generated fixed event. In 72% of games the text of the event was corrected. The total number of changes (of both entities or choice events) was 120. The changes were necessary especially for the wrong choices, with 54 total changes (Fig. 4).

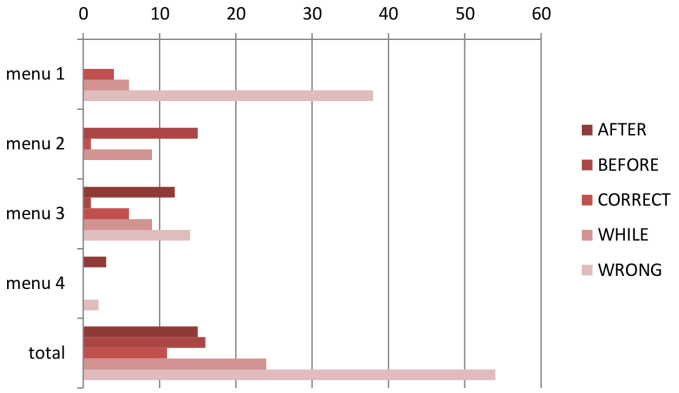

Fig. 4 Distribution of changes

The work of developing manual causal games was instead longer, as the operator had to invent the game by directly using, when possible, one of the fixed events already present in the story games. Overall, 75 causal games were created (both cause, effect, and cause/effect games). The average time spent for their development was equal to 23 minutes per game. The total work, also including game loading and final review, was about 30 minutes for each causal game.

3 Conclusion

Most of the revisions were of a formal nature, e.g. correct the letters for the names, verbs tenses. The main effort was in connection with reviewing the text: the phrases generated were incomplete or inconsistent, so it was necessary to continue to work on accents, the verb tenses and sentence length. A quantitative revision analysis showed a good level of automatic generation: in only 15 cases it was necessary to change the fixed automatically generated event. Mostly wrong type solutions were changed, with 54 change made necessary primarily to ensure a consistent level of difficulty in finding this solution.

Therefore, important and crucial improvements shall be directed towards the annotations of causal links (absent in this release), the text generation task, and the heuristic that takes care of the distractors (i.e., the wrong choices, and especially when the annotations are poor).

Acknowledgements This work was supported by the TERENCE project, funded by the EC through FP7 for RTD, ICT-2009.4.2.
Special thanks to the team: A. Carbonelli, M.R. Cecilia, G. Cofini, G. D'Ascanio, M. Palmerio, J. Ranieri, M. Santucci, G. Sementilli, F. Zazzara.

References

1. Cofini, V., Di Giacomo, D., Di Mascio, T., Vittorini, P.: Evaluation Plan of TERENCE: when the User-centred Design Meets the Evidence-based Approach. In: Proc. of the ebTEL 2012 workshop co-located with PAAMS 2012. Springer (2012)
2. Cofini, V., de la Prieta, F., Di Mascio, T., Gennari, R., Vittorini, P.: Design Smart Games with requirements, generate them with a Click, and revise them with a GUIs. Advances in Distributed Computing and Artificial Intelligence Journal 1(3), 59–68 (2012)
3. Cofini, V., Di Mascio, T., Gennari, R., Vittorini, P.: The TERENCE smart games revision tool. In: Proc. of the ebTEL 2013 workshop co-located with PAAMS 2013. Springer (2013)
4. De La Prieta Pintado, F., Di Mascio, T., Gennari, R., Marenzi, I., Vittorini, P.: Playing for Improving the Reading Comprehension Skills of Primary School Poor Comprehenders. In: 1st International Workshop on Pedagogically-driven Serious Games (2012-09-01)
5. De La Prieta Pintado, F., Di Mascio, T., Gennari, R., Marenzi, I., Vittorini, P.: The TERENCE Smart Games: Automatic Generation and Supporting Architecture. In: 1st International Workshop on Pedagogically-driven Serious Games (2012-09-01)
6. Gennari, R.: Deliverable 4.1: M.State of the Art and Design of Novel Intelligent Feedback Deliverable 4.1. Tech. rep., TERENCE project (2011)
7. Gennari, R.: Deliverable 4.2: Automated Reasoning Module. Tech. Rep. D4.2, TERENCE project (2012)
8. Hedberg, S.R.: Executive Insight: Smart Games: Beyond the Deep Blue Horizon. IEEE Expert 12(4), 15–18 (1997)
9. Moens, S.: Deliverable 3.1: State of the Art and Design of Novel Annotation Languages and Technologies. Tech. Rep. D3.1, TERENCE project (2012)
10. TERENCE Consortium: TERENCE web site, http://www.terenceproject.eu