An Agent Based Competitive Translation Game for Second Language Learning

Wang Ling, Isabel Trancoso, Rui Prada

INESC-ID / Instituto Superior Técnico, Portugal

wang.ling@inesc-id.pt, isabel.trancoso@inesc-id.pt, rui.prada@ist.utl.pt

Abstract

This paper describes a competitive language translation game aimed at improving students vocabulary and writing skills. An automated agent is employed as an opponent in order to improve the user's motivation and maintain the user focused. The agent's actions are based on statistical machine translation outputs. An evaluation that was performed with 20 Portuguese learners of Mandarin suggested that users were more focused and motivated when playing against the agent than playing alone. Furthermore, the majority of students felt that the system helped them learn Mandarin and would like to use it in the future. The system has a web-based implementation and is easily accessible by language learners.

Index Terms: computer assisted language learning, machine translation, automated agent, educational games

1. Introduction

In many educational games, automated agents have been successfully used to improve their effectiveness [1]. However, the majority of computer assisted language learning (CALL) systems use the agents in a very limited form. Most often, the agent serves as a simple mediator that is used to assign tasks to the users and give feedback on their answers. This is the case of the "translation game" [2][3], in which the user is asked to translate a sentence in his/her native language into a target language, receiving feedback on its correctness. Learning and perfecting a foreign language requires long periods of practice, and is largely affected by the students motivation and their capacity to stay focused. However, motivating students and making them stay focused for large periods of time is not simple, specially for children. In our work, we propose a Competitive version of the "translation game", where players play against an Automated Agent. We believe that playing against an opponent will make the players more motivated and focused during their tasks, improving the effectiveness of the game. Such an approach was shown to be effective in a "programming course" [4]. The factor that motivated the students was the challenge provided by their automatic opponent, which encouraged them to spend more time on their tasks and revising their code, in order to win.

Our game is also centered around the task of translating a sentence from the players native language into the language the player wishes to learn. The rules of our system, however, are different from the original "translation game", since they were not suited for a competitive game. Although in the original game, it is possible to compare the correctness of the student's answer with the opponent's answer, the actions from one player do not influence the other. Furthermore, the correctness is evaluated by translating the source sentence using machine translation systems, and assuming that the agent would also use such a system to give the answer, it would be very hard to beat the agent, even if the translation models differ. In our game, one word is proposed by each player at each time, which allows players some strategies, such as answering easier words first and leaving harder ones to the opponent. Rather than testing whether the player can convey the same meaning in his/her translation, testing whether the answer is similar enough to the a correct translation, our game will be more strict, only allowing a single translation, selected among a set of manually created references for that sentence. Also, rather than teaching the user how to speak fluently, our game's goal is to expand the user's vocabulary and syntatic skills.

2. System Overview

Users of the system can choose among a set of predefined exercises and whether they want to play by themselves or against an opponent. Furthermore, they can choose to take the interactive tutorial, where they can learn the basics of the game. All students that used in this game were asked to fill the optional survey at the end, which is described in section 3. The game is available at "http://www.l2f.inesc-id.pt/TranslationGame/?mode=main&lang=en"(Tested with the Firefox Browser 3.6.16).

2.1. Rules

Each game is composed by a number of rounds. In each round the system presents a sentence in the user's source language (typically, the user's native language), and the corresponding sentence in the target language with a number of hidden words (or characters, in the Chinese case), marked with an empty underlined space. Players take turns to guess the words that are hidden, and a player can only propose one word at each turn. Players are rewarded 20 points when they get the right answer and penalized 5 points when they propose a wrong answer. They can also choose to pass their turn if they do not know the answer to avoid losing points. When no correct answer is given after 4 turns (e.g. when both players pass twice), a hidden word is revealed. This is done to handle the case where both players do not know the answer. In each round, the hardest word to find is marked in yellow, which is worth 40 points. Finally, the player who guesses the last word completing the sentence receives an additional 30 points. This is because players will guess the words they know first and leave the words that they are unsure about for last, which makes the last word more likely to be harder. The round ends when all words are filled. After all rounds end, the game terminates and the winner is the player who scored more points.

2.2. Interface

Figure 1 is an illustration of the game interface. The source sentence is marked with the word "A" and the translation is marked with the word "B". In the translation, the words are also indexed

by the number above the word so that they can be easily referenced. Words that are guessed by the user are colored in green and words that are guessed by the opponent are colored in red. The words that are colored black are either words that are initially revealed or revealed when multiple turns passed without any correct word proposal. Hidden words such as the word at index 4 are left empty with a blue underline. To propose an answer for a hidden word, players can click on it and a prompt will pop up where users can propose write in the word. Afterwards, a message will appear for a few seconds on the top left corner of the screen, alerting the user if the proposed word was correct or not. The system also gives feedback on the opponents actions in the same way. To pass a turn, users can click on the "pass turn" button in the panel marked by "E". The players' scores are displayed in the player information panels "C" and "D". Finally, players can mouse over a hidden word to see a pop up with all incorrect words that were already proposed.

2.3. Exercise Creation

The exercises were created by processing the BTEC and DIALOG test corpus from the IWSLT evaluation [5], which are parallel corpora in the language pairs French-English and English-Chinese, respectively. Each sentence in the source language has 16 references in the target language, which are manual translations done by different translators. In order to lower the ambiguity in possible translations we use the available references and for each word we calculate the number of references it appears in, normalized by the number of references, which we will call word ambiguity. Words that appear in all references tend to be less ambiguous. For instance, in the sentence "J'aimerais contacter l'ambassade Japonaise", which had 16 references, the words "I", "Japanese" and "Embassy" appear in mostly all references, while ambiguos words such as "want", was freely changed with "like", "wish", "need". After that, we select the reference with the least average in word ambiguity for all words in that reference. This reference will be henceforth designated as $t_{correct}$. Then, we select one third of the words in that sentence as hint words, giving higher priority to words with higher word ambiguity. Finally, from the remaining hidden words, we choose the one that has the lowest word ambiguity as the bonus word.

2.4. Agent

A simple method to create an automatic opponent would be to set it to pass turns and insert a random word from translation every few turns. However, this type of agents would not simulate a human like behavior, which would lead to a poor user experience. Human players would for instance propose incorrect answers, miss harder words more often and try to avoid answering words that would give hints to the other player (for instances phrasal verbs, such as "pass away" would be much easier to guess after one of the words is known). While it is possible to simulate some of the functions above, such as generating similar words to simulate wrong word proposals, doing most of these tasks is not trivial.

In order to create a human-like opponent behavior we define an utility based agent which uses machine translation systems to find the answers. The agent has a representation of its current state S, and a Markov process that determines how that state evolves for each action a the agent or the player can perform. A state S includes features such as whether it is the agent's or the user's turn, the words that are already found or revealed, the scores of each player and the incorrect words that were proposed for each hidden word. There are two types of actions that the agent can perform: to pass the turn or to propose a word. The pass action is deterministic, since we know that the next state S^{t+1} will have the same features, except that it will be the opponent's turn. If the proposed word is correct, the next state S_a^{t+1} can have one less hidden word. If it is incorrect, the next state S_b^{t+1} will have the same hidden words, and the opponent will know that the answer the agent gave is not the correct one for that index. We define the next set of possible states for each action as $NS(a)$. The Markovian process $T(S^{t+1}, a, S)$ determines the probability of the state S evolving into the state S^{t+1} after action a. With these definitions, there are two important results that need to be determined. First, we need to determine the possible proposal type actions for a given state and the probability of the proposal being correct. Secondly, we need to define an algorithm to find the best action for the current state.

In every state, the pass action is always in the set of possible actions. The possible word proposals are generated using the output of a statistical machine translation (SMT) system, fed with the source language sentence. Since the most probable translation from the SMT system might not be the correct answer, we generate our word proposals from the list of 5000 best translations t_x from the SMT system, ordered by correctness, where x is the position of the translation in the list. We estimate the probability of a word in the translation t_x at index i being the correct word for the hidden word at index j in the translation $t_{correct}$ using a linear combination of 5 weighted features. First, we give higher probabilities for words that occur higher in the translation list. Secondly, we measure the correctness of the translation t_x, compared to the $t_{correct}$. This is done by measuring how many of the words in t_x are found in the revealed words in $t_{correct}$, counter-measured by how many the already given incorrect words are in t_x. The third feature calculates the distance between i and j. The fourth feature penalizes the proposal of words that are already in the $t_{correct}$ sentence. Finally, we also introduce a noise feature, which has a random value, which can be used to lower the difficulty of the agent, by setting it higher. Presently, it is set to 0.

To determine the best action from the current state we use the minimax algorithm. This decision algorithm allows the agent to choose the action that minimizes the possible loss while maximizing the potential gain. In our game, the gain of one participant is the loss of the other (zero-sum game). Thus, the agent will choose the action that maximizes its gain, while the user will try to choose actions that will minimize the gain of the agent. $BestAction(S)$ is hence defined as the action a that maximizes the gain $G(a)$ if it is the agent's turn, and the action that minimizes the gain, if it is the user's turn. Since actions can be non-deterministic, we adopt the approach presented in [6], to adapt the minimax algorithm for a non-deterministic algorithm and define the gain $G(a)$ as:

$$G(a) = \sum_{S^{t+1} \in NS(a)} T(S^{t+1}, a, S) \times U(S^{t+1}) \quad (1)$$

This equation means that the gain of an action is the sum of the utilities $U(S)$ over all possible next states, weighted by the probability of going to that state. The utility of each state is given by:

$$U(S) = \begin{cases} ScoreDifference(S) & if\ FinalState(S) \vee DepthReached(d) \\ G(BestAction(S)) & otherwise \end{cases} \quad (2)$$

The utility of a state depends on the action that can be performed on that state by the following player, and not by the score difference on that state. For instance, even if we score a word,

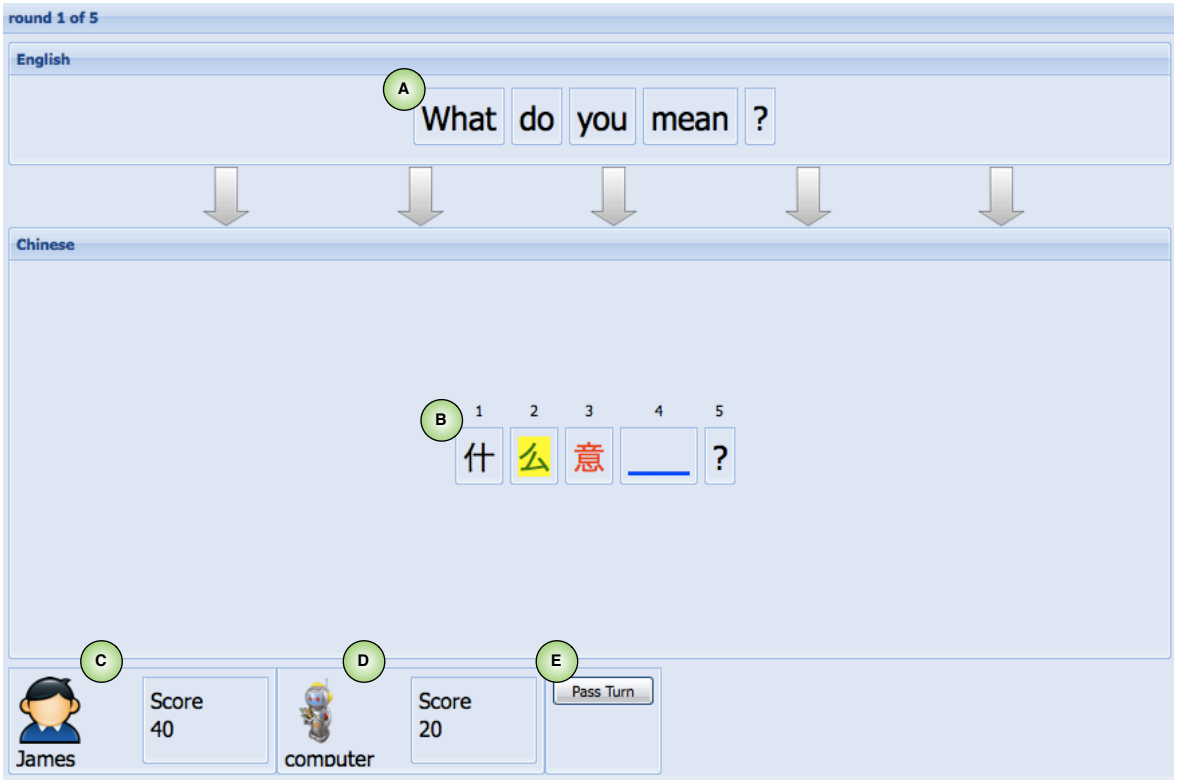

Figure 1: *Web interface for the translation game. 'A' is the sentence in the source language, 'B' is the sentence in the target language, 'C' is the information panel of the human player, 'D' is the information panel of the target player, and 'E' is the button panel.*

raising our points, that state might have a low utility if our word makes it easier for the the next player to guess another word, specially if that word is worth more points. The only exception is when the game ends with the agent's action, which means that $FinalState(S) = TRUE$, in which case there is nothing the other player can do, if a depth d is reached $DepthReached(d)$. The depth can be seen as the number of turns the agent can think ahead, and it is used because of time and computational limitations. In our game, let the algorithm run for 5 seconds at most to avoid the degeneration of the game experience due to inactivity.

3. Preliminary Evaluation

To validate our claims and to evaluate the system we conducted a user study with Portuguese speaking students learning Mandarin. The test was conducted in the Missão Macau (http://www.portugalvirtual.pt/0/5527dat1.html) facilities in Lisbon and in the Centro Científico e Cultural de Macau (http://www.cccm.mctes.pt/), where weekly Mandarin classes are given. The students are distributed upon entrance in different levels raging from 1 to 5, where 5 is the most proficient level. The classes are taught by the same professors in both locations. The system was tested with 9 students from MM (Missão Macau) from levels 4 and 5, and 11 students from CCCM (Centro Científico e Cultural de Macau) from levels 2 and 3. We did not use level 1 students since the tests were performed on a period when the student's vocabulary was still very limited.

Participants were first taught how to use the system with an interactive tutorial (http://www.l2f.inesc-id.pt/TranslationGame/?mode=tutorial&lang=en). The test conductors illustrated how to write Chinese characters in a computer, for those who did not know. Each participant played two sets of 5 exercises, in different setups. In the first setup, the agent was not present, while in the second it was. The set of exercises was empirically chosen to include different levels of exercises so that both sets were equivalent in difficulty. The test is available at (http://www.l2f.inesc-id.pt/TranslationGame/?mode=survey&lang=en). The adaptation of the game to the language pair Portuguese-Chinese was done by manually translating the English source sentences to Portuguese in the two sets of exercises, for the purpose of this evaluation.

Afterwards, the participants were asked to fill a survey composed by 7 closed questions, with answers in the Likert scale. The answers ranged from "I completely agree", "I agree", "I don't agree, nor disagree", "I disagree" and "I completely disagree". These are valued from 1 to 5, in the same order. The English version of the questions is given below:

- Q1: Learning how to use the system was easy.
- Q2: The exercises were useful for learning Mandarin.
- Q3: The choice of the bonus word was adequate.
- Q4: The challenge provided by the automatic opponent was adequate.
- Q5: I was more motivated when playing against an automatic opponent.
- Q6: I was more focused when playing against an automatic opponent.
- Q7: I would use this game as a helping tool to learn Mandarin.

Partipants were also asked their ages, gender and whether they used computers regularly.

The results for all participants are shown in table 1 for each question. This table contains the average (AVG) and standard deviation (STDDEV), with multiple participant groupings. First, we group all 20 partipants together (ALL). Then, we grouped the students by age (Age $>$ 35, and Age $<=$ 35). Finally, we separated the students by level (Levels 2 & 3 vs. Levels 4 & 5).

Table 1: *Survey Results*

Results	Q1	Q2	Q3	Q4	Q5	Q6	Q7
ALL (20)							
AVG	1.75	1.65	2.2	2.15	1.9	1.85	1.8
STDDEV	0.55	0.75	0.62	0.59	0.91	0.81	0.62
Age $\leq$ 35 (9)							
AVG	1.89	1.22	2.56	2	1.33	1.44	1.78
STDDEV	0.33	0.44	0.73	0	0.5	0.73	0.44
Age $>$ 35 (11)							
AVG	1.64	2	1.91	2.27	2.36	2.18	1.81
STDDEV	0.67	0.77	0.30	0.79	0.92	0.75	0.75
Levels 2 & 3 (11)							
AVG	1.91	1.27	2.45	2	1.45	1.55	1.73
STDDEV	0.30	0.47	0.69	0	0.52	0.69	0.47
Levels 4 & 5(9)							
AVG	1.56	2.11	1.89	2.33	2.44	2.22	1.89
STDDEV	0.73	0.78	0.33	0.87	1.01	0.83	0.78

The participant ages ranged from 10 to 84. At CCCM, the students were younger, with an average of 28.1, ranging from 10 to 58, whereas the participants from MM had an average of 58.0, ranging from 43 to 84. 8 partipants were female and 12 were male.

Overall the responses were very positive. The means of Q2 and Q7 regarding the usefulness and willingness to use the tool were both below 2. Most participants found the system intuitive to use, after the tutorial, but there were some cases of older participants that had some difficulty using the system. As a result, the average score for Q1 was also low (1.75).

We noticed that younger participants felt more motivated and focused when playing against an opponent, with an average score for Q5 and Q6 of 1.33 and 1.44, respectively, vs. an average score of 2.36 and 2.18 for older participants. A t-test showed that the statistical significance of the answers for Q5 (p-value=0.008) was higher than the answer for Q6 (p-value=0.04). The same can be observed when clustering participants by level. We believe that lower level students felt more motivated and focused when playing against an opponent because the level of the agent during the exercises is equal or higher than the level of the participants. From our observation during the experiment, the participants from levels 2 and 3 scores were on par with the opponent scores, which made them more motivated and focused. On the other hand, the gap between the scores of the participants from levels 4 and 5 and the opponent was much bigger, even after only a few rounds. Thus, they felt less challenged by the opponent except in the first rounds, when the score was still tied. In this case, p-values for the answers for Q5 and Q6 were 0.011 and 0.062, respectively. The average score for participants from levels 2 and 3 on Q4, which asks whether the challenge from the opponent was appropriate, was slightly lower than the answers from participants from levels 4 and 5. While the statistical significance of this result is low (p-value=0.215), this is suggests that participants that lost heavily against the opponent felt that the opponent's level should be adjusted to the player, which we were not able to do since we had no prior knowledge of the participants skills and the number of exercises that were performed were not enough to find a good estimate of the level of the participant during the test.

4. Conclusions and Future Work

In this paper we have described a Web-based competitive translation game, intended to help students learn a second language. Players can play against an intelligent agent that uses statistical machine translation output to perform the tasks in the game. The evaluation that was performed with 20 students learning Mandarin led us to believe that, in general, the participants felt more motivated and focused when playing against the agent. It also revealed that the players that felt more motivated and focused were the students whose levels were on par with the agent. On the other hand, students with levels worse than the agents also felt more motivated and focused, but thought that the level of the agent was not adequate. This suggests that, as future work, a mechanism should be developed to progressively adjust the level of the agent to the level of the student. The extension of the game to encompass a much broader set of exercises or other language pairs is straightforward, since it is limited in finding enough parallel corpora. The institution where the tests were performed will be providing the digital books used in their lessons. These books are divided by level of proficiency and contain the set of Mandarin characters on taught on each level. This will allow us to cluster the exercises by level and we will perform a long term evaluation of the progress of the students in each level using our system.

5. Acknowledgements

This work was partially supported by FCT (INESC-ID multiannual funding) through the PIDDAC Program funds, and also through projects CMU-PT/HuMach/0039/2008 and CMU-PT/0005/2007. The PhD thesis of Wang Ling is supported by FCT grant SFRH/BD/51157/2010. The authors would like to thank all the participants present in the evaluation and their professors without whom this work would not be possible.

6. References

[1] C. Conati and M. Klawe, "Socially intelligent agents to improve the effectiveness of educational games," 2000.

[2] C. Wang and S. Seneff, "A spoken translation game for second language learning," in *Proceeding of the 2007 conference on Artificial Intelligence in Education: Building Technology Rich Learning Contexts That Work*. Amsterdam, The Netherlands, The Netherlands: IOS Press, 2007, pp. 315–322.

[3] M. Rayner, P. Bouillon, N. Tsourakis, J. Gerlach, M. Georgescul, Y. Nakao, and C. Baur, "A multilingual call game based on speech translation." in *LREC'10*, 2010, pp. –1–1.

[4] R. Lawrence, "Teaching data structures using competitive games," *Education, IEEE Transactions on*, vol. 47, no. 4, pp. 459 – 466, nov. 2004.

[5] M. Paul, M. Federico, and S. Stüker, "Overview of the iwslt 2010 evaluation campaign," in *IWSLT '10: International Workshop on Spoken Language Translation*, 2010, pp. 3–27.

[6] S. Koenig and R. G. Simmons, "Real-time search in non-deterministic domains," in *Proceedings of the 14th international joint conference on Artificial intelligence - Volume 2*. San Francisco, CA, USA: Morgan Kaufmann Publishers Inc., 1995, pp. 1660–1667.