Teaching iSTART to Understand Spanish

Mihai Dascalu[1,2], Matthew E. Jacovina[2], Christian M. Soto[3], Laura K. Allen[2], Jianmin Dai[2], Tricia A. Guerrero[2], & Danielle S. McNamara[2]
[1] University "Politehnica" of Bucharest, Faculty of Automatic Control and Computers, 313 Splaiul Independenței, RO-60042 Bucharest, Romania
mihai.dascalu@cs.pub.ro

[2] Arizona State University, Institute for the Science of Teaching and Learning, PO Box 8721111, Tempe, AZ 85287
{matthew.jacovina, laura.k.allen, jianmin.dai, taguerre, dsmcnama}@asu.edu
[3] University of Concepción, Psycholinguistics Area, Department of Spanish, Victor Lamas 1290, Concepción, Chile
christiansoto@udec.cl

Dascalu, M., Jacovina, M. E., Soto, C., Allen, L. K., Guerrero, T. & McNamara, D. S. (2017). Teaching iSTART to understand Spanish. In R. Baker & E. Andre (Eds.), *Proceedings of the 18th International Conference on Artificial Intelligence in Education* (pp. 485-489), Wuhan, China: Springer. Published with acknowledgment of federal support.

Author's Note

This work was partially funded by the FP7 2008-212578 LTfLL project, by University Politehnica of Bucharest through the "Excellence Research Grants" Program UPB–GEX 12/26.09.2016, as well as by the Institute for the Science of Teaching & Learning (IES R305A130124) and the Office of Naval Research (ONR N000141410343 and ONR N00014-17-1-2300).

Teaching iSTART to Understand Spanish

Mihai Dascalu[1,2], Matthew E. Jacovina[2], Christian M. Soto[3], Laura K. Allen[2], Jianmin Dai[2], Tricia A. Guerrero[2], & Danielle S. McNamara[2]

[1] University "Politehnica" of Bucharest, Faculty of Automatic Control and Computers, 313 Splaiul Independenței, RO-60042 Bucharest, Romania
mihai.dascalu@cs.pub.ro

[2] Arizona State University, Institute for the Science of Teaching and Learning, PO Box 8721111, Tempe, AZ 85287
{matthew.jacovina, laura.k.allen, jianmin.dai, taguerre, dsmcnama}@asu.edu

[3] University of Concepción, Psycholinguistics Area, Department of Spanish, Victor Lamas 1290, Concepción, Chile
christiansoto@udec.cl

Abstract. iSTART is a web-based reading comprehension tutor. A recent translation of iSTART from English to Spanish has made the system available to a new audience. In this paper, we outline several challenges that arose during the development process, specifically focusing on the algorithms that drive the feedback. Several iSTART activities encourage students to use comprehension strategies to generate self-explanations in response to challenging texts. Unsurprisingly, analyzing responses in a new language required many changes, such as implementing Spanish natural language processing tools and rebuilding lists of regular expressions used to flag responses. We also describe our use of an algorithm inspired from genetics to optimize the Fischer Discriminant Function Analysis coefficients used to determine self-explanation scores.

Keywords: Reading comprehension, Natural Language Processing, Intelligent Tutoring Systems, Optimizing score prediction.

1 Introduction

Intelligent tutoring systems (ITS) provide students with automated instruction and feedback based on their performance and current level of understanding [1]. Providing this adaptive instruction can present a number of challenges for developers, particularly for ITSs that target ill-defined domains such as reading comprehension and writing. In these domains, assessment is highly complex and depends on myriad factors related to the individual student, the task properties, and the specific types of errors that are made. One assessment technique that has been used in these domains is Natural Language Processing (NLP) techniques. In particular, NLP has been used to analyze students' written responses as a means of measuring their performance across multiple domains, such as writing ability and reading comprehension [2, 3, 4].

Despite their success, these algorithms are limited in their ability to generalize to multiple languages. This poses problems, as quality reading comprehension instruction is needed for students across multiple languages. In this paper, we describe the process of adapting a web-based reading comprehension tutor, iSTART, such that it is fully available in Spanish. In particular, we describe the process of refining the NLP algorithm that guides scoring and feedback in the system, and outline several of the techniques that we considered and used throughout this translation process.

iSTART (Interactive Strategy Training for Active Reading and Thinking) is a web-based ITS that provides adolescent students with instruction and practice on reading comprehension strategies. In particular, iSTART focuses on strategies for generating high quality *self-explanations* during the reading process [5]. An algorithm scores the quality of these self-explanations on a scale from 0 (response is too short or nonsensical) to 3 (response makes global connections, often bringing in information from beyond the text) in terms of how well students are using the self-explanation strategies [3]. This scoring algorithm relies on a combination of word-based measures and Latent Semantic Analysis (LSA) to provide scores on each generated self-explanation [3].

2 iSTART-E: iSTART en Español

There is increasing evidence that students in Spanish-speaking countries struggle to demonstrate proficiency on standardized assessments of reading comprehension and writing skills [6]. In response to this issue, several educational technologies have been developed to target Spanish literacy [7, 8], and have been shown to improve students' performance on reading comprehension tasks [9]. However, much more work is needed to better promote ITSs in Spanish-speaking countries, including empirical examinations of the similarities and differences in these systems across English- and Spanish-speaking populations.

In light of this goal, we have worked to develop a Spanish version of the iSTART system – iSTART-E [10]. Similar to any large translation project, there were numerous important decision points, ranging from how to translate the instructional materials to be understandable for the widest range of Spanish speakers, to the selection of practice texts that would be contextually meaningful for many Spanish speakers. We outline in Table 1 important steps that had to be made to develop iSTART-E, as well as the specific techniques used to implement these changes.

In addition, we implemented an evolutionary algorithm [11] based on metaheuristics inspired from genetics, more specifically bio-inspired processes such as selection, mutation and crossover in order to improve the accuracy of iSTART-E's scoring algorithm. In general, genetic algorithms provide a high degree of flexibility in generating high-quality solutions to search and optimization problems: hence, their applicability to our task. The purpose of the designed algorithm was to determine the optimal Discriminant Function Analysis (DFA) coefficients that create the best mapping between the scores of Spanish self-explanation translations and those already computed for English. The previously mentioned coefficients represent the Fischer classification coefficients corresponding to the three discriminant functions that are currently used in the automated iSTART English scoring mechanism [3].

Table 1. Encountered challenges and provided solutions.

Challenge	Implemented solution
Providing a new dictionary of words	Upon manual review, we selected the dictionary found at http://www.winedt.org/dict.html which includes low-frequency, scientific words
Introducing a new list of stop words	The stop words list from Snowball (http://snowball.tartarus.org/algorithms/spanish/stop.txt) was expanded to include words describing numbers and interjections (e.g., "bah")
Correcting misspelled words	Instead of the Soundex algorithm available for English, we implemented a rule matching algorithm that relies on the Levenshtein edit distance
Tagging important words from practice texts	The Stanford Core NLP for Spanish (http://stanfordnlp.github.io/CoreNLP) was integrated and used to identify content words (i.e., nouns, verbs, adjectives, adverbs)
Normalizing words to allow for comparisons	Our static lemmatizer based on predefined transformation http://www.lexiconista.com/datasets/lemmatization/ automatically changes word forms to their corresponding inflectional form, i.e. lemma
Building an LSA space in Spanish	The LSA space was built using Apache Mahout with the lemmatized corpus provided by El Grupo de Interés en el Análisis de la Semántica Latente (http://elsemantico.es/index.html)
Translating regular expressions	The algorithm uses regular expressions to identify various types of self-explanations and to flag special types of responses. Manual corrections were made due to language specificities
Iterative testing	We manually translated 2,982 English self-explanations from the identification mini-games into Spanish. Distributions of the various components captured by the algorithms (e.g., matching content words; LSA cosine values) and repeated measures ANOVAs for each component across the two languages helped us identify discrepancies

3 Results and Discussion

At the start of the development of the iSTART-E Spanish algorithm, we calculated agreement between the English and Spanish algorithms on a preliminary set of 588 self-explanations (i.e., a subset of the self-explanations that were used to tune the scoring algorithm using the genetic algorithm). The exact agreement was 43.46% and the adjacent agreement was 84.35%. These agreement scores are a baseline, albeit imperfect, to compare our agreement results following the techniques and solutions described in this paper along with other bug fixes and refinements.

A total of 1,638 out of the 2,982 self-explanations, that were manually translated from English, were not flagged to receive special feedback and were included in the follow-up analyses. Before using the new (genetic) algorithm, but after implementing the changes described in Table 1, the exact accuracy was 64.90% with 89.87% adjacent accuracy. After applying the evolutionary algorithm with all considered optimizations, we obtained an exact accuracy of 69.5% with 94.10% adjacent accuracy, an increase in terms of aligning Spanish self-explanations to the corresponding English scores.

Although the correspondence between the Spanish and English scoring algorithms has improved from the initial accuracy, several concerns still exist. Broadly, we note that it will never be possible to perfectly align the iSTART system across the two

languages – particularly with respect to the scoring algorithms. For instance, our scoring systems must rely on different LSA spaces that consist of completely different text sources. Moreover, refining the pre-processing stage for Spanish also creates potential differences between the algorithms.

In sum, the optimization of a scoring algorithm for a different language poses many challenges, but is not insurmountable. The complete transformation of a system such as iSTART into a new language requires attention to detail and sufficient input from numerous sources that are sensitive to changes in the language.

Acknowledgements. This work was partially funded by the FP7 2008-212578 LTfLL project, by University Politehnica of Bucharest through the "Excellence Research Grants" Program UPB–GEX 12/26.09.2016, as well as by the Institute for the Science of Teaching & Learning (IES R305A130124) and the Office of Naval Research (ONR N000141410343 and ONR N00014-17-1-2300).

References

1. VanLehn, K.: The behavior of tutoring systems. International Journal of Artificial Intelligence in Education, 16, 227–265 (2006)
2. Boonthum, C., McCarthy, P.M., Lamkin, T., Jackson, G.T., Magliano, J., McNamara, D.S.: Automatic natural language processing and the detection of reading skills and reading comprehension. In: Proceedings of FLAIRS, AAAI Press, Palm Beach, FL (2011)
3. McNamara, D.S., Boonthum, C., Levinstein, I.B., Millis, K.: Evaluating self-explanations in iSTART: Comparing word-based and LSA algorithms. In: Landauer, T.K., McNamara, D.S., Dennis, S., Kintsch, W. (eds.) Handbook of Latent Semantic Analysis, pp. 227–241. Erlbaum, Mahwah, NJ (2007)
4. Allen, L.K., Snow, E.L., McNamara, D.S.: Are you reading my mind? Modeling students' reading comprehension skills with natural language processing techniques. In: 5th Int. Learning Analytics & Knowledge Conf. (LAK'15), pp. 246–254. ACM, Poughkeepsie, NY (2015)
5. Jackson, G.T., McNamara, D.S.: Motivation and performance in a game-based intelligent tutoring system. Journal of Educational Psychology, (105), 1036–1049 (2013)
6. Programme for International Student Assessment (PISA): PISA 2015 Results in Focus. OEDC Publishing (2016)
7. Vidal-Abarca, E., Gilabert, R., Ferrer, A., Ávila, V., Martínez, T., Mañá, A., Llorens, A.C., Gil, L., Cerdán, R., Ramos, L., Serrano, M.A.: TuinLEC, an intelligent tutoring system to improve reading literacy skills / TuinLEC, un tutor inteligente para mejorar la competencia lectora. Infancia y Aprendizaje, 37, 25–56 (2014)
8. Véliz, M., Osorio, J.: Desarrollo de un software para el desarrollo de la capacidad de lectura crítica. RLA: Revista de lingüística teórica y aplicada, 39, 203–220 (2001)
9. Ponce, H.R., López, M.J., Mayer, R.E.: Instructional effectiveness of a computer-supported program for teaching reading comprehension strategies. Computers & Education, 59(4), 1170–1183 (2012)
10. Soto, C.M., McNamara, D.S., Jacovina, M.E., Snow, E.L., Dai, J., Allen, L.K., Perret, C.A., Johnson, A.M., Russell, D.G.: iSTART-E: Desarrollando un tutor inteligente para la comprensión lectora de estudiantes de habla hispana. In: M. García (Ed.), Proceedings of FLAIRS. AAAI Press, Orlando, FL (2015)
11. Mitchell, M.: An Introduction to Genetic Algorithms. MIT Press, Cambridge, MA (1996)