

The TREACLE-ALEGRO timeline: from error-annotated corpus to adaptive learning tool

Oksana Polyakova^{1*}, Penny MacDonald²

¹Universidad Católica de Valencia

²Universitat Politècnica de València

*Corresponding author's e-mail: oksana.polyakova@ucv.es

Abstract

In recent years, much research has been devoted to the error analysis but few have improved the visualization and correction of errors. In this report, we describe the study of error-annotated corpus which embodies some basic principles of error coding. TREACLE-ALEGRO, our recent projects for development of an adaptive learning tool, are the solution to a few challenges as such EFL grammar teaching curriculum and web-based language learning system.

1 Introduction

Corpus linguistics as a reliable language analysis tool has proved to be an effective approach towards natural sampling research. In addition, the possibility to extract teaching samples and exercises from the corpus to target learners' errors at different proficiency levels has been our goal for nearly a decade.

This report includes such research questions as: What are the most frequent errors made by Spanish university students? How can we transfer them to personalised teachable items?

Motivated by these premises, a group of Spanish researchers and university lecturers started a cooperative error-annotation project back in 2010. It was an important point to begin dealing with the real needs of Spanish students who aim to improve English grammar skills. The project is now being continued with the operational development of an adaptive learning tool that includes the most significant grammar rules detected.

The rest of the report is organised as follows. Section 2 provides a brief picture of the project outcomes and methodology applied. We place our work in context of two stages: the TREACLE project explores the main issues concerning corpus analysis and error coding, meanwhile the ALEGRO project moves the prior results to a new level of an online learning tool development. Finally, Section 3 draws some conclusions and recommendations for further actions.

2 Projects overview and methodology

2.1 THE TREACLE PROJECT (2010-2013)

As we already mentioned, the research has two broad aims and, therefore, two specific purposes. On one hand, it is the grammar analysis of students' corpora by looking at the errors learners make (stage 1). On the other, it involves identifying the mistakes and elaborating training procedures (stage 2).

The first stage of the study is related to the TREACLE project (*Teaching Resource Extraction from an Annotated Corpus of Learner English*) funded by the Spanish Government (Ministerio de Ciencia e Innovación, FFI2009-14436/FILO).

The error-coded corpus contained the following documents:
Essays written in English by Spanish university students.
304 texts (110,000 words) providing 16,000 errors

Source: 78 texts (67,000 words) from WriCLE corpus (Rollinson & Mendikoetxea, 2010);

226 texts (44,000 words) from the MiLC corpus (Andreu et al., 2010).

Quick Oxford Placement test scores taken at time of writing helped detect proficiency levels (UCLES 2001).

According to Murcia-Bielsa and MacDonald (2013), major results after automatic syntactic tagging and manual error annotation done by using UAM Corpus Tool (O'Donnell 2008) are the following:

- error areas: grammatical, lexical, punctuation, pragmatic, and phrasing;
- relevant grammar categories: np-error, prep-phrase-error, vp-error, clause-error (see Figure 1 below):

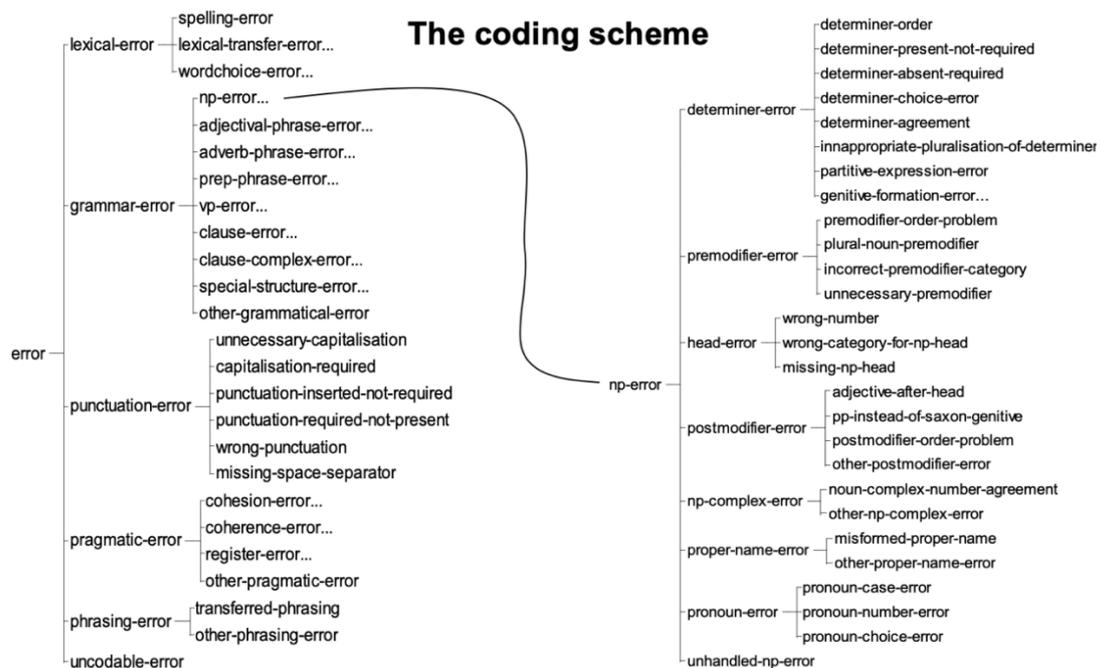


FIGURE 1 The TREACLE coding scheme

- most frequent grammar errors: determiner-present-not-required, prep-choice-error, subject-finite-agreement, determiner-absent-required, wrong-number (see Figure 2 below):

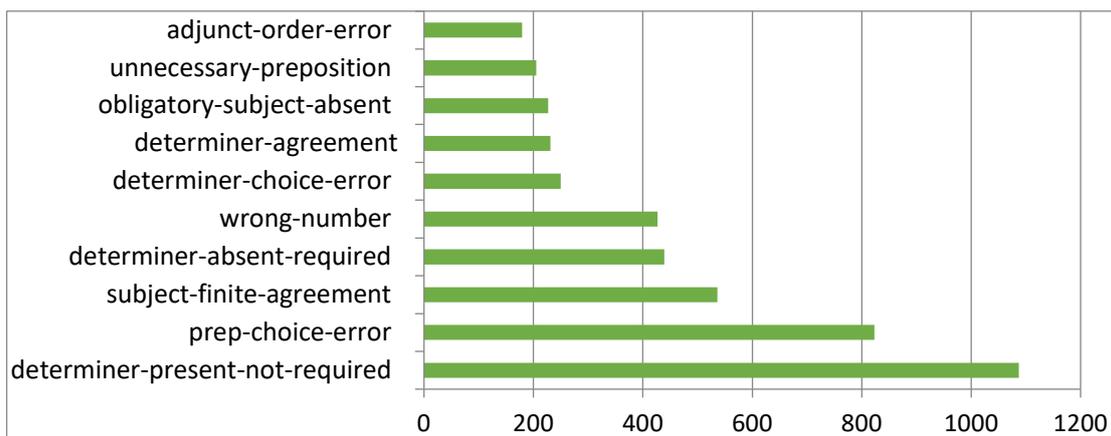


FIGURE 2 10 most frequent grammar errors in TREACLE

As a result, the TREACLE project approached English learner proficiency from two perspectives: correct and incorrect use of grammar structures (Murcia-Bielsa & MacDonald 2013). Furthermore, the stage 1 of the timeline intends for automatic grammatical tagging of the corpus to be initial part of the process. The major focus is on manual error annotation that will provide clear connections between grammar mastery at each proficiency level as well as a list particular structures university students are still acquiring/ learning.

2.2 THE ALEGRO PROJECT (ONGOING)

The second phase of the study is connected to the ALEGRO project (*Adaptive Learning of English Grammar Online*) funded by the Spanish Government (Ministerio de

Economía y Competitividad, FFI2015-67992-R).

The goal of our current project is to develop an online learning system assisting Spanish university students in the acquisition of important grammatical concepts. In order to achieve it, an in-depth error exploration (MacDonald 2016), error coarse-graining and teachable concepts elaboration (O'Donnell 2018, O'Donnell et al. 2018) were required.

The major advantage of the ALEGRO approach consists in its adaptation to learner's need. The online training system tracks learner assimilation of concepts and offers grammatical themes to study that each learner requires individually. For example, as Figure 3 shows, a student is presented with a short explanation of concepts and several examples of correct use and broken rules. After that, he/she can take quizzes on the concepts.



FIGURE 3 ALEGRO content editor

4 Conclusions

In our research, we proved that a combined linguistic and online technology based approach can be used to target English language mistakes and the TREACLE-ALEGRO projects are no exception to it. Our methodology for the error analysis and development of an adaptive learning tool helps redesign second language acquisition curriculum by motivating students to undertake a personalised training.

References

- [1] Andreu M, Astor A, Boquera M, MacDonald P, Montero B, Pérez C 2010 Analysing EFL learner output in the MiLC project: An error *it's, but which tag? In M.C. Campoy, B. Belles-Fortuno & M.L. Gea-Valor (eds.) *Corpus-Based Approaches to English Language Teaching*. London: Continuum 167-79
- [2] Hawkins J A, Buttery P 2009 Using learner language from corpora to profile levels of proficiency: Insights from the English Profile Programme. In L. Taylor & C. J. Weir (eds.) *Language Testing Matters: Investigating the wider social and educational impact of assessment* Cambridge: Cambridge University Press 158-75
- [3] Rollinson P, Mendikoetxea A 2010 Learner corpora and second language acquisition: Introducing WriCLE. In J. L. Bueno Alonso, D. González Álvarez, U. Kirsten Torrado, A. E. Martínez Insua, J. Pérez-Guerra, E. Rama Martínez & R. Rodríguez Vázquez (eds.), *Analizar datos>Describir variación / Analysing data>Describing variation*. Vigo: Universidade de Vigo (Servizo de Publicacións) 1-12
- [4] O'Donnell M 2008 The UAM CorpusTool: Software for corpus annotation and exploration In Bretones Callejas, Carmen M. et al. (eds) *Applied Linguistics Now: Understanding Language and Mind / La Lingüística Aplicada Hoy: Comprendiendo el Lenguaje y la Mente*. Almería: Universidad de Almería 1433-47
- [5] O'Donnell M 2012 Using learner corpora to redesign university-level EFL

Acknowledgments

The TREACLE project, <http://www.treacle.es>:

- Susana Murcia, Mick O'Donnell, Clara Molina, Rebeca García, Esther Mediero, Ainhoa Robles;
- Penny MacDonald, Keith Stuart, María Boquera, Ana Botella, Laura Cardona.

The ALEGRO Project, <http://www.alegro.org.es>:

- Mick O'Donnell, Susana Murcia, Clara Molina, Patricia González, Esther Mediero, Fiorella Dotti.
- Penny MacDonald, Begoña Montero, Begoña Clavel, Oksana Polyakova.

- [6] O'Donnell M 2018 Towards targeted online grammar learning: The Alegro project. Invited keynote at the Jornadas de Innovación Anglotic 20-21 November, 2018
- [7] O'Donnell M, Polyakova O, MacDonald P 2018 Applying an error corpus to the classroom: from coarse-grained errors to teachable concepts *Paper presented at the X Congreso Internacional de Lingüística de Corpus, Cáceres, Spain, 9-11 May, 2018*
- [8] MacDonald P 2016 'We al make mistakes!' Analysing an Error-coded Corpus of Spanish University Students' Written English *Complutense Journal of English Studies* 24 103-29
- [9] Murcia-Bielsa S, MacDonald P 2013 The TREACLE project: Profiling learner proficiency using error and syntactic analysis In S. Granger, G. Gilquin and F. Meunier (eds.) *Twenty Years of Learner Corpus Research: Looking back, Moving ahead. Corpora and Language in Use – Proceedings 1*. Louvain-la-Neuve: Presses universitaires de Louvain 335-44 ISBN 978-2-87558-199-0
- [10] UCLES 2001 *Quick Placement Test (Paper and pencil version)*. Oxford: Oxford University Press