Exploring Errors by EFL Students at University Level

ASYRAS (6th-8th October 2011)

Authors:
Ainhoa Robles Baena
M.ª Esther Mediero Durán

Funded by the Ministry of Science and Innovation (FFI2009-14436/FILO)
I. Introduction & Main Aim
II. What is TREACLE?
   I. Purpose of a Computer-Aided Error Annotation
III. The Corpus
IV. Methodology & Taxonomy
V. Problems and Solutions
VI. Results
VII. Conclusions
I. Introduction

Aim of paper:

- Analyse university students’ written production at different levels of linguistic competence.
- Using an error coding scheme to identify most problematic areas for students at every level.
Aims at:

- Producing grammatical profiles of Spanish University students' written English Language.

- Linking relevant areas to the CEFR levels with the aim of redesigning the English Language grammar curriculum to improve its efficiency and projection.

*For more information, visit: [http://www.uam.es/treacle](http://www.uam.es/treacle)
### III. The Corpus ➔ MiLC + Wricle

<table>
<thead>
<tr>
<th></th>
<th>Amount of Texts</th>
<th>Length of texts</th>
<th>Genre / Topic</th>
<th>Students</th>
<th>Level</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>MiLC Corpus</strong></td>
<td>780 texts ➔ 150,000 words</td>
<td>220-250 words per text</td>
<td>Mainly devoted to Immigration.</td>
<td>Spanish University Students of all Levels from UPV</td>
<td>A1-C1</td>
</tr>
<tr>
<td>– UPV (Andreu et al 2010)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Wricle Corpus</strong></td>
<td>521 essays ➔ 500,000 words</td>
<td>1,000 avg. words per text</td>
<td>Essays / Variety of topics such as immigration, homosexual marriages, traffic problems.</td>
<td>Students of English Philology</td>
<td>A2 - C2 as per CEFR</td>
</tr>
<tr>
<td>– UAM (Rollinson &amp; Mendikoetzxea 2010)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
IV. Methodology

- WHAT DO WE ANALYSE?: (*Error vs mistakes*)

- HOW DO WE ANALYSE IT?
  - *Manual Annotation with UAM Corpus Tool (O'Donnell, 2008)*.

- ERROR CODING PROCESS
- ERROR TAXONOMY
- PROBLEMS AND SOLUTIONS

Error-coding Process

1. Select text containing error.
2. Provide the corrected text here.
3. Assign features to current segment here.
The TREACLE project: Error Taxonomy (1)

- Error scheme devised by the research team
  - Ability to change the scheme as needed
  - Glosses to help coders
  - Contains 113 errors at most delicate level
The TREACLE project:
Error Taxonomy (2)

- Five main error types + uncodable error:
  - Lexical errors
  - Grammatical errors
  - Punctuation errors
  - Pragmatic errors
  - Phrasing errors
Error Taxonomy

- error
  - lexical-error
    - lexical-transfer-error...
    - wordchoice-error...
  - grammar-error
    - np-error...
    - adjectival-phrase-error...
    - adverb-phrase-error...
    - prep-phrase-error...
  - vp-error...
    - clause-error...
    - clause-complex-error...
    - special-structure-error...
    - other-grammatical-error
  - error
    - unnecessary-capitalisation
    - capitalisation-required
    - punctuation-inserted-not-required
    - punctuation-required-not-present
    - wrong-punctuation
    - missing-space-separator
  - punctuation-error
    - cohesion-error...
    - coherence-error...
    - register-error...
    - other-pragmatic-error
  - phrasing-error
    - transferred-phrasing
    - other-phrasing-error
  - uncodable-error
V. Problems & Solutions

- **PROBLEMS**
  - Many Coders
  - Coding problems:
    - segmentation & error code

- **SOLUTIONS**
  - 2 ICRS – Inter-Coder Reliability Studies with the aim of:
    - Refining the “*Error Scheme*” and the “*Coding Criteria Manual*”.
    - Ensuring all coders were complying with the coding criteria.

- Consensus Rates: Reliability scores calculated by comparing each individual's coding to the consensus model.
Pedagogically Oriented Segmentation.

When segmenting, the following rules apply:

1. Minimal segmentation
2. Correct what has been written and not what should have been written.
The TREACCLE project: Example 1 / ICRS Segmentation

The education in Spain is a subject that given a lot of play because for one people this system of education is great, but for other people is awful.'

ven a lot of play because for one people this system of education is great, but for other people is awful.'

<table>
<thead>
<tr>
<th>Consensus: grammar-error: np-error: determiner-error: determiner-choice-error</th>
<th>some</th>
</tr>
</thead>
<tbody>
<tr>
<td>✓ R7</td>
<td>✓</td>
</tr>
<tr>
<td>✓ R1</td>
<td>✓</td>
</tr>
<tr>
<td>✓ R2</td>
<td>✓</td>
</tr>
<tr>
<td>✓ R5</td>
<td>✓</td>
</tr>
<tr>
<td>✗ R4 grammar-error: np-error: premodifier-error: incorrect-premodifier-category</td>
<td>✓</td>
</tr>
<tr>
<td>✗ R3 lexical-error: wordchoice-error: other-wordchoice-error</td>
<td>✓</td>
</tr>
</tbody>
</table>

ven a lot of play because for one people this system of education is great, but for other people is awful.'

| R6 grammar-error: np-error: determiner-error: determiner-choice-error | some people |
The TREACLE project: Example 2 ICRS Error code assigned

‘there are a lot of players who have an important paper’.

- Disagreements like this were solved during the consensus meetings.

<table>
<thead>
<tr>
<th>Consensus: lexical-error: wordchoice-error: other-wordchoice-error: noun-vocab-error</th>
<th>role</th>
</tr>
</thead>
<tbody>
<tr>
<td>✓ R4</td>
<td>✓</td>
</tr>
<tr>
<td>✓ R1</td>
<td>✓</td>
</tr>
<tr>
<td>✗ R7 phrasing-error: transferred-phrasing</td>
<td>✓</td>
</tr>
<tr>
<td>✗ R6 lexical-error: wordchoice-error: transferred-word: borrowing</td>
<td>✓</td>
</tr>
<tr>
<td>✗ R3 lexical-error: wordchoice-error: false-friend</td>
<td>✓</td>
</tr>
<tr>
<td>✗ R2 lexical-error: wordchoice-error</td>
<td>✓</td>
</tr>
</tbody>
</table>
VI. Results so far

**Coded Texts:** 146 (63,000 words)

**Errors:** 8,021
Results per Level

- Errors per CEFR Levels
  - B1 – Nearly 50% grammatical errors.
  - C2 – Punctuation and pragmatic errors.
Results: Most salient categories

Types of Errors:

44% Grammar Errors

→ Noun Phrase: → Determiner (30%)

→ Prepositions
Error Categories per Level

![Error Categories Graph]

- np-error
- adjectival-phrase-error
- adverb-phrase-error
- prep-phrase-error
- vp-error
- clause-error
- clause-complex-error

Legend:
- a1
- a2
- b1
- b2
- c1
- c2
VII. Conclusions

- The error-coding system we are using is viable and seems to provide codes for almost all of the errors we have encountered.
- The number of errors decreases as the proficiency level increases.
- More emphasis is needed in the teaching of NP (noun phrase), syntax and prepositions.
- Although time-consuming, this kind of work is extremely useful to provide data on learners' interlingua and contribute to a potential improvement in the design of the curriculum.
- This type of studies is essential for a more fine-grained grammar syllabus at the university level.
Thank you

Ainhoa Robles (ainhoarobles@hotmail.com)
Esther Mediero (medieroesther@hotmail.com)
Web page TREACLE: http://www.uam.es/treacle