Lexical Errors and Language Acquisition by University Level EFL Students

Author: M.ª Esther Mediero Durán

Funded by Ministerio de Ciencia e Innovación (FFI2009-14436/FILO)
Outline

1. Introduction & Main Aim
2. The Project
3. Methodology
4. Error Scheme
5. Results
6. Conclusions
1 Introduction

Aim of this paper:

- Analyse university students’ written production at different levels of linguistic competence.
- Explore the lexical errors they make to gain a better insight into the more problematic areas.
2.1 The Project: TREACLE

- Project: TREACLE
- Teaching Resource Extraction from an Annotated Corpus of Learner English
- A cooperation between: Universidad Autónoma de Madrid and Universitat Politècnica de Valencia
- Funded by Ministerio de Ciencia e Innovación (FFI2009-14436/FILO)
- Runs: January 2010 – June 2013

Official Title: “Developing an annotated corpus of learner English for pedagogical application”
2.2 Goals of the Project

- Use learner English corpora to **profile** the lexical and grammatical skills of Spanish university learners at each proficiency level (A1, A2, B1, etc.)

- Use these profiles to **redesign the teaching curriculum**: determining which lexical & grammatical features need to be taught/ reinforced, in what order, and with what degree of emphasis.

- Provide a **web-based language learning system** which dynamically adapts to the student.

- More information available at [http://www.uam.es/treacle](http://www.uam.es/treacle)
2.3 The Corpora

- The project uses two corpora:
  - The **WriCLE** corpus (UAM) - *Written Corpus of Learner English*. 521 essays of ~1000 words each, written by Spanish learners of English at University level (about 500,000 words) (Rollinson and Mendikoetxea 2008)
  - The **UPV Learner Corpus** (UPV) containing 150,000 words of shorter texts by ESP students. (Andreu et al 2010)

- Quick Oxford Placement test (UCLES, 2001) given at same time, to measure proficiency

- Other metadata: gender, academic year, degree, parent languages, time abroad, resources used in writing, etc.
3. Methodology

- ERROR CODING PROCESS

- PROBLEMS & SOLUTIONS
- CODING CRITERIA
3.1 Error-coding Process

1. Select text containing error.

2. Provide the corrected text here.

3. Assign features to current segment here.
3.2 Problems & Solutions

RELIABILITY:

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

- Some examples.
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 educati

<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</td>
<td>grammar-error: np-error: premodifier-error: incorrect-premodifier-category</td>
</tr>
<tr>
<td>✖ R3</td>
<td>lexical-error: wordchoice-error: other-wordchoice-error</td>
</tr>
</tbody>
</table>

ven a lot of play because for one people this system of education is g

| R6 | grammar-error: np-error: determiner-error: determiner-choice-error | some people |
When segmenting, the following rules apply:
1. Minimal segmentation
2. Correct what has been written and not what should have been written.
Example 2 below shows some of the doubts the coders had initially as regards the exact identification of the error type: ‘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:</th>
<th>role</th>
</tr>
</thead>
<tbody>
<tr>
<td>lexical-error: wordchoice-error: other-wordchoice-error: noun-vocab-error</td>
<td></td>
</tr>
<tr>
<td>✓ R4</td>
<td>✓</td>
</tr>
<tr>
<td>✓ R1</td>
<td>✓</td>
</tr>
<tr>
<td>✓ R7</td>
<td>✓</td>
</tr>
<tr>
<td>✗ R6</td>
<td>✓</td>
</tr>
<tr>
<td>✗ R3</td>
<td>✓</td>
</tr>
<tr>
<td>✗ R2</td>
<td>✓</td>
</tr>
</tbody>
</table>
4. Error Scheme - The TREACLE Project

- 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
  - Coded by 8 coders/members.
  - 2 Inter-coder reliability studies
  - Post-coding Revision: currently undergoing.
4.1 Error Taxonomy

- **Lexical Error**
  - Lexical Transfer Error
  - Lexical Intra-Language Error
  - NP Error
  - Adjectival Phrase Error
  - Adverb Phrase Error
  - Prep Phrase Error
  - VP Error
  - Clause Error
  - Clause Complex Error
  - Special Structure Error
  - Morphological Error
  - Other Grammatical Error
  - Unnecessary Capitalisation
  - Capitalisation Required
  - Punctuation Inserted Not Required
  - Punctuation Required Not Present
  - Wrong Punctuation
  - Space Separator Error

- **Grammatical Error**
  - Grammatical Unit

- **Punctuation Error**
  - Punctuation Error Type
  - Cohesion Error
  - Coherence Error
  - Register Error
  - Other Pragmatic Error
  - Other Punctuation Error

- **Pragmatic Error**
  - Pragmatic Error Type

- **Phrasing Error**
  - Phrasing Error Type
  - Other Phrasing Error

- **Uncodable Error**
4.2 The error scheme – Main types

- Five main error types + uncodable error:
  - Lexical errors
  - Grammatical errors
  - Punctuation errors
  - Pragmatic errors
  - Phrasing errors
4.3 Lexical Errors

- lexical-error
  - lexical-transfer-error
    - transfer-where-similar-exists
    - false-friend
    - malformation
  - word-creation
    - coinage
    - borrowing
  - dictionary-error
    - spelling-error
    - wordchoice-error
  - lexical-intralanguage-error
5 Results

- Errors per level
- General Trend
- Main error categories
- Lexical errors per category
Errors Made per Level

Errors / 1,000 words

- a1
- a2
- b1
- b2
- c1
- c2
Results per Level

- **Total Errors:** 16,000 (110,000 words)
- **Errors per CEFR Levels**
  - A1 – More lexical errors
  - B1 – Nearly 50% grammatical errors.
  - C2 – Punctuation and pragmatic errors.
Results per Level: Main Error Categories

- **a1**: 46% (grammar-error), 23% (lexical-error), 8% (punctuation-error), 7% (pragmatic-error)
- **a2**: 46% (grammar-error), 23% (lexical-error), 8% (punctuation-error), 7% (pragmatic-error)
- **b1**: 49% (grammar-error), 23% (lexical-error), 8% (punctuation-error), 7% (pragmatic-error)
- **b2**: 42% (grammar-error), 23% (lexical-error), 8% (punctuation-error), 7% (pragmatic-error)
- **c1**: 39% (grammar-error), 23% (lexical-error), 8% (punctuation-error), 7% (pragmatic-error)
- **c2**: 34% (grammar-error), 23% (lexical-error), 8% (punctuation-error), 7% (pragmatic-error)
Types of Errors per Level

- noun-based-lex-error
- verb-based-lex-error
- adverb-based-lex-error
- adjective-based-lex-error
- other-wc-lex-error
Types of lexical errors per level

- lexical-transfer-error
- lexical-intralanguage-error
Lexical Errors in terms of apparent difficulty

- malformation
- coinage
- false-friend
- transferred-spelling
- verb-vocab-error
- spelling-error
- adverb-vocab-error
- borrowing
- noun-vocab-error
- adjective-vocab-error

More common with basic learners

More common with advanced learners
Lexical Errors in terms of apparent difficulty

With the exception of borrowing, Transfer errors are more common for beginners, while later, intralanguage errors predominate.

Borrowings at advanced levels: more explicit mention of Spanish institutional terms: “Fiscal Jefe”

More common with basic learners

Malformation
Coinage
False-friend
Transferred-spelling
Verb-vocab-error
Spelling-error
Adverb-vocab-error
Borrowing
Noun-vocab-error
Adjective-vocab-error

More common with advanced learners
<table>
<thead>
<tr>
<th>Transfer errors</th>
<th>Intralingual Errors</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Borrowing</strong></td>
<td><strong>Coinage</strong></td>
</tr>
<tr>
<td>carril-bici</td>
<td>determined</td>
</tr>
<tr>
<td>laboral</td>
<td>optative</td>
</tr>
<tr>
<td>España</td>
<td>fomenting</td>
</tr>
<tr>
<td>ONGs</td>
<td>course (verb)</td>
</tr>
<tr>
<td>Europa</td>
<td>sanity (health)</td>
</tr>
<tr>
<td>temporal</td>
<td>poblation</td>
</tr>
<tr>
<td>mas</td>
<td>form</td>
</tr>
<tr>
<td>hachis</td>
<td>displacements</td>
</tr>
<tr>
<td>mundial</td>
<td>asignature</td>
</tr>
<tr>
<td>conducta</td>
<td>desesperation</td>
</tr>
<tr>
<td>infantil</td>
<td>diary</td>
</tr>
<tr>
<td>habituate</td>
<td>principately</td>
</tr>
<tr>
<td>receptor</td>
<td>evolution</td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Conclusions of this study

- Lexical errors do fall as learners progress, both in terms of frequency of occurrence (per 1000 words) and in terms of the proportion of errors.

- Within lexical errors, the nontransfer errors dominate at all levels, however, the proportion of errors due to transfer is clearly falling with increased proficiency.

- In terms of lexical errors and word class, there is a movement away from noun-based errors towards more errors with verbs. Adjective based errors also increase at lower levels but then stabilise.
Thank you

Esther Mediero (medieroeesther@hotmail.com)
Web page TREACLE: http://www.uam.es/treacle