



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)

Outline



- 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.*

II. TREACLE* (*Teaching Resource Extraction from an Annotated Corpus of Learner English*)



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>

III. The Corpus → MiLC + Wricle



| | Amount of Texts | Length of texts | Genre / Topic | Students | Level |
|--|----------------------------|---------------------------|---|--|---------------------|
| MiLC Corpus – UPV (Andreu et al 2010) | 780 texts → 150,000 words | 220-250 words per text | Mainly devoted to Immigration. | Spanish University Students of all Levels from UPV | A1- C1 |
| Wricle Corpus – UAM (Rollinson & Mendikoetxea 2010) | 521 essays → 500,000 words | 1,000 avg. words per text | Essays / Variety of topics such as immigration, homosexual marriages, traffic problems. | Students of English Philology | A2 - C2 as per CEFR |

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**

*Free software available at <http://www.wagsoft.com/CorpusTool/>

Error-coding Process



Error analysis for: Files/A101-4.txt

The legalisation of Marijuana has become a very polemic topic in our present society because it concerns a lot of different opinions and because the decision of legalizing it or not will directly affect a whole society. So, legalizing Marijuana is a social debate whose outcome couldn't possibly please everyone. In this essay I will try to discuss the different perspectives and opinions of soft-drugs such as Marijuana.

The general opinion of people who argue that if the government legalize this type of drugs the crime rates will increase and many youths will be able to obtain the drugs without any problem at all. They also believe that drugs are very dangerous substances that can cause enormous problems to the

1. Select text containing error.

Navigation: << < > >> Ignore Delete Other Action... Save Close Help

| Assigned | ERROR-TYPE | Gloss |
|---------------|-------------------------|-------|
| error | np-error | |
| grammar-error | adjectival-phrase-error | |
| | adverb-phrase-error | |
| | prep-phrase-error | |
| | vp-error | |
| | clause-error | |
| | clause-complex-error | |

Correction: the

Comment:

2. Provide the corrected text here.

3. Assign features to current segment here.

The TREACLE project: Error Taxonomy (1)



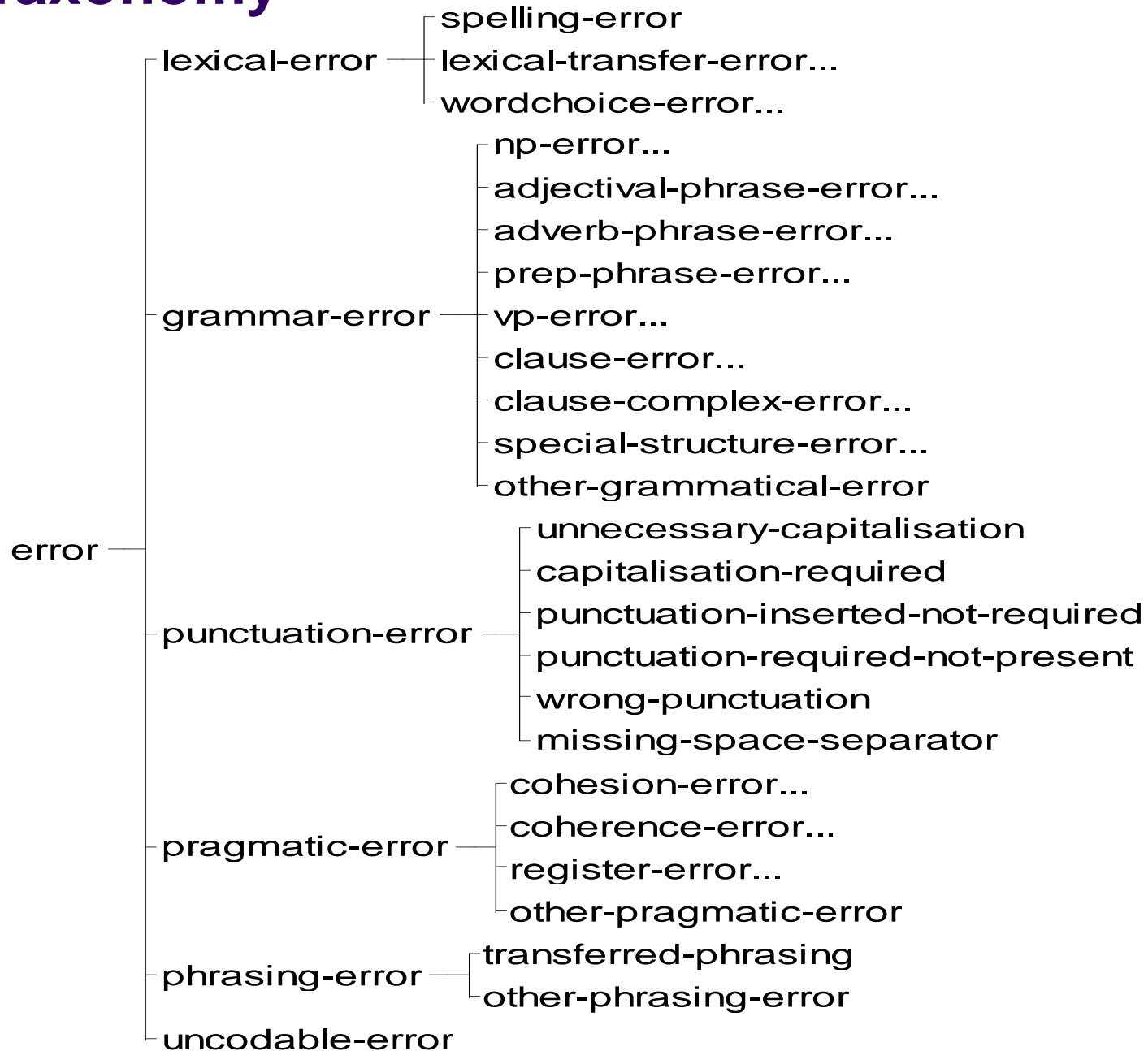
- Error scheme devised by the research team
 - Ability to change the scheme as needed
 - Glosses to help coders
 - Contains 113 errors at most delicate level

The TREACLE project: Error Taxonomy (2)



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

Error Taxonomy



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.**

The TREACLE project: Example 1 - ICRS Segmentation



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 TREACLE 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

| | | | |
|---|------------|--|------|
| | Consensus: | grammar-error: np-error: determiner-error: determiner-choice-error | some |
| ✓ | R7 | | ✓ |
| ✓ | R1 | | ✓ |
| ✓ | R2 | | ✓ |
| ✓ | R5 | | ✓ |
| ✗ | R4 | grammar-error: np-error: premodifier-error: incorrect-premodifier-category | ✓ |
| ✗ | R3 | lexical-error: wordchoice-error: other-wordchoice-error | ✓ |

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 |
|---|----|--|-------------|

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.

players who have an important **paper**. In this aspect I have to men

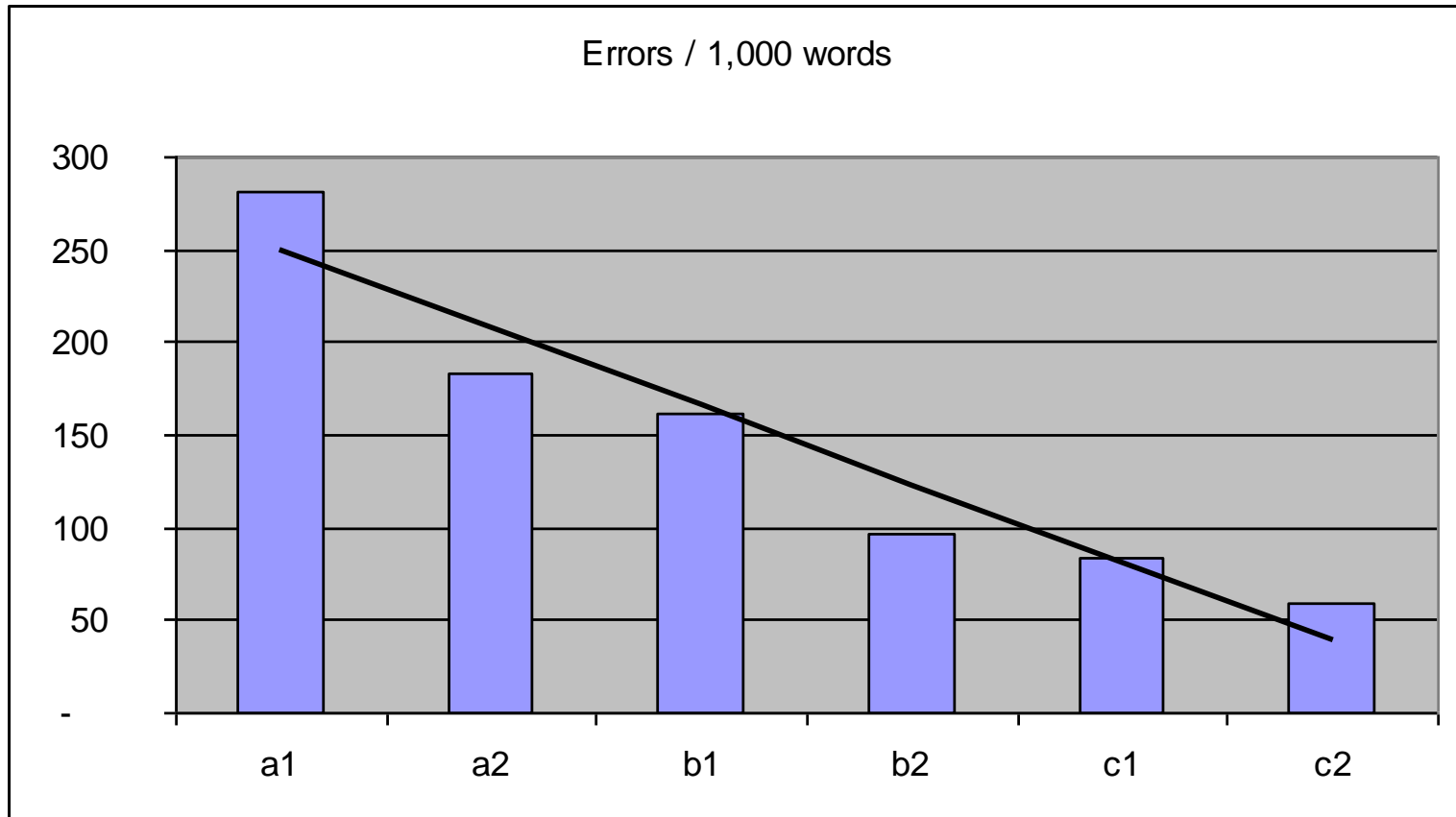
| | Consensus: | lexical-error: wordchoice-error: other-wordchoice-error: noun-vocab-error | role |
|---|------------|---|------|
| ✓ | R4 | | ✓ |
| ✓ | R1 | | ✓ |
| ✗ | R7 | phrasing-error: transferred-phrasing | ✓ |
| ✗ | R6 | lexical-error: wordchoice-error: transferred-word: borrowing | ✓ |
| ✗ | R3 | lexical-error: wordchoice-error: false-friend | ✓ |
| ✗ | R2 | lexical-error: wordchoice-error | ✓ |

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



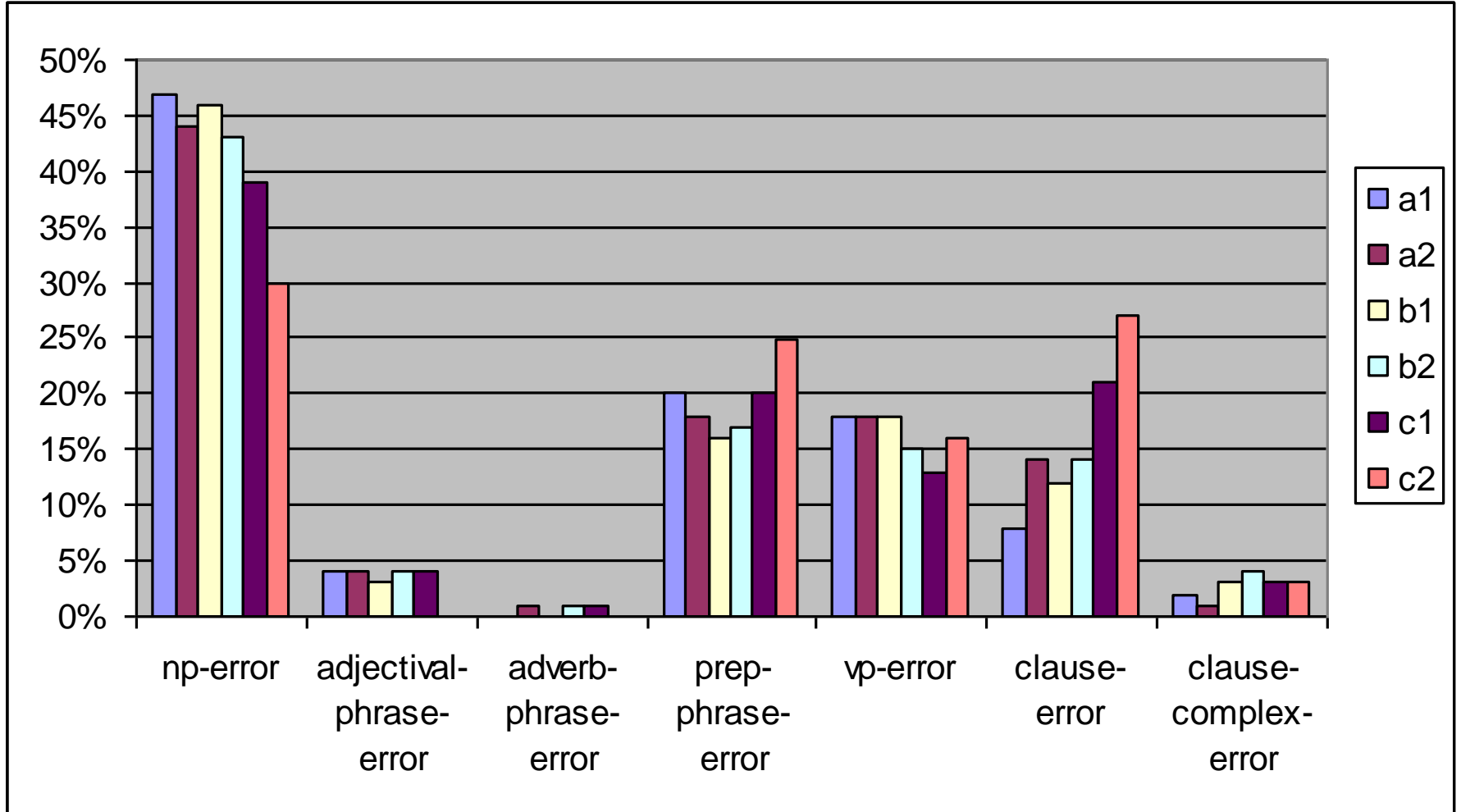
Types of Errors:

44% Grammar Errors

→ Noun Phrase: → Determiner (30%)

→ Prepositions

Error Categories per Level



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>