

# **Lexical Errors and Language Acquisition**

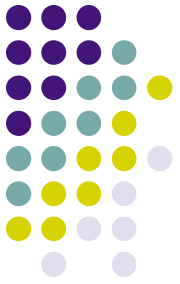
by University Level EFL Students

Author:

M.<sup>a</sup> Esther Mediero Durán

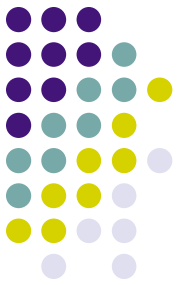


# Outline



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



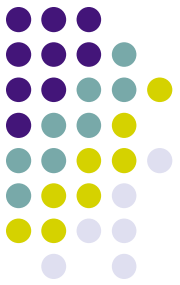
- Project: TREACLE



T  
R  
E  
A  
C  
L  
E  
Teaching  
Resource  
Extraction from an  
Annotated  
Corpus of  
Learner  
English

*Official Title: “Developing an annotated corpus of learner English for pedagogical application”*

- 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



## 2.2 Goals of the Project

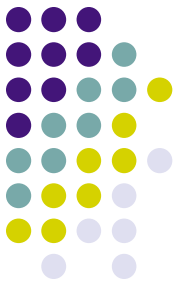
- Use learner English corpora to **profile** the lexical and grammatical skills of Spanish university learners at each proficiency level (A1, A2, B1, etc.)
- Use these profiles to **redesign the teaching curriculum**: determining which lexical & grammatical features need to be taught/ reinforced, in 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>

## 2.3 The Corpora



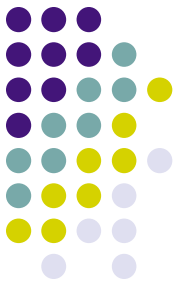
- The project uses two corpora:
  - ☞ The **WriCLE** corpus (UAM) - *Written Corpus of Learner English*. 521 essays of ~1000 words each, written by Spanish learners of English at University level (about 500,000 words)  
(Rollinson and Mendikoetxea 2008)
  - ☞ The **UPV Learner Corpus** (UPV) containing 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**
  - Manual Annotation with UAM Corpus Tool (O'Donnell, 2008).
- **PROBLEMS & SOLUTIONS**
- **CODING CRITERIA**

# 3.1 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 are against the legalization of drugs is 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.

<< < > >> 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	

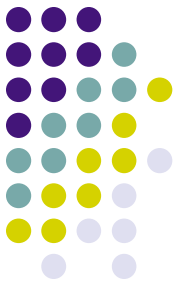
3. Assign features to current segment here.

Correction: the

Comment:

2. Provide the corrected text here.

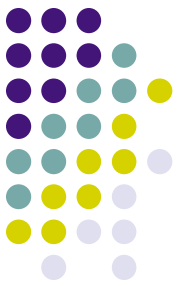




## 3.2 Problems & Solutions

### 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 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

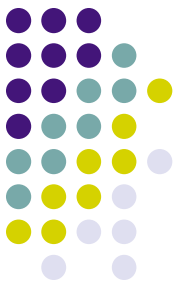
✗	R6	grammar-error: np-error: determiner-error: determiner-choice-error	some people
---	----	--	-------------

# The TREACLE Project - Example 1: ICRS 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 2 - ICRS Error code assigned

**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.

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

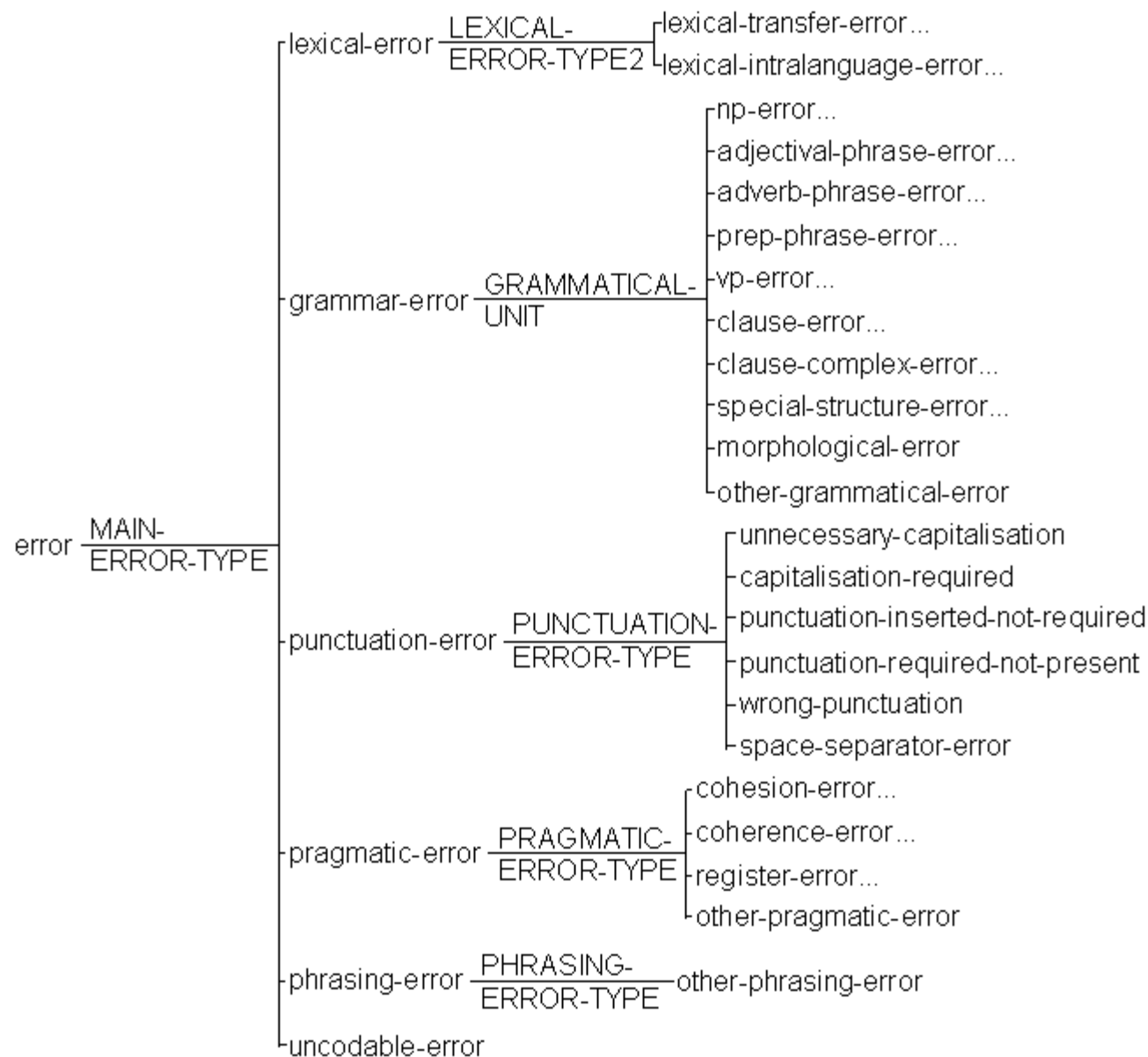
	Consensus:	lexical-error: wordchoice-error: other-wordchoice-error: noun-vocab-error	role
✓	R4		✓
✓	R1		✓
✗	R7	phrasing-error: transferred-phrasing	✓
✗	R6	lexical-error: wordchoice-error: transferred-word: borrowing	✓
✗	R3	lexical-error: wordchoice-error: false-friend	✓
✗	R2	lexical-error: wordchoice-error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



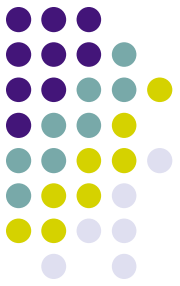
## 4. Error Scheme - The TREACLE Project

- 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

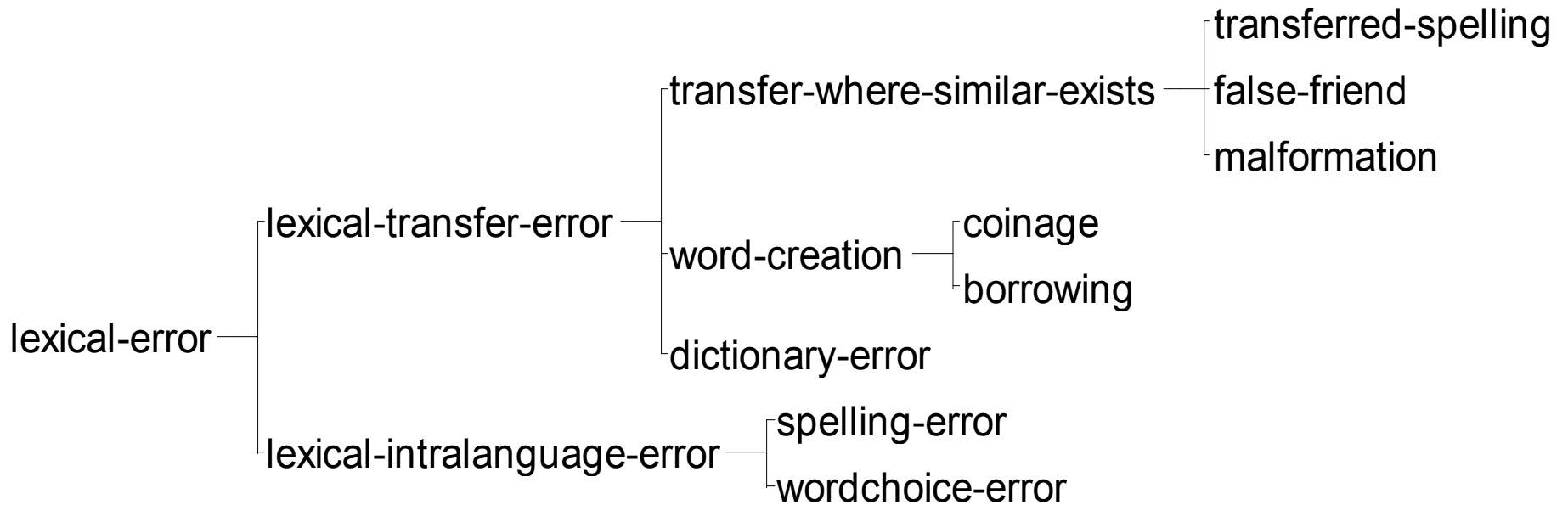
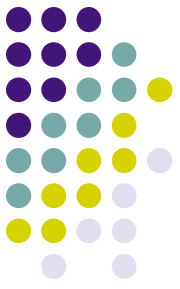


## 4.2 The error scheme – Main types



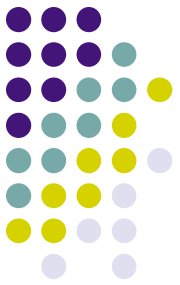
- Five main error types + uncodable error:
  - Lexical errors
  - Grammatical errors
  - Punctuation errors
  - Pragmatic errors
  - Phrasing errors

## 4.3 Lexical Errors



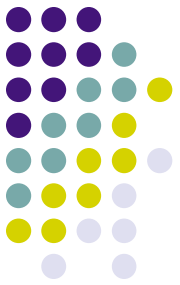


# 5 Results

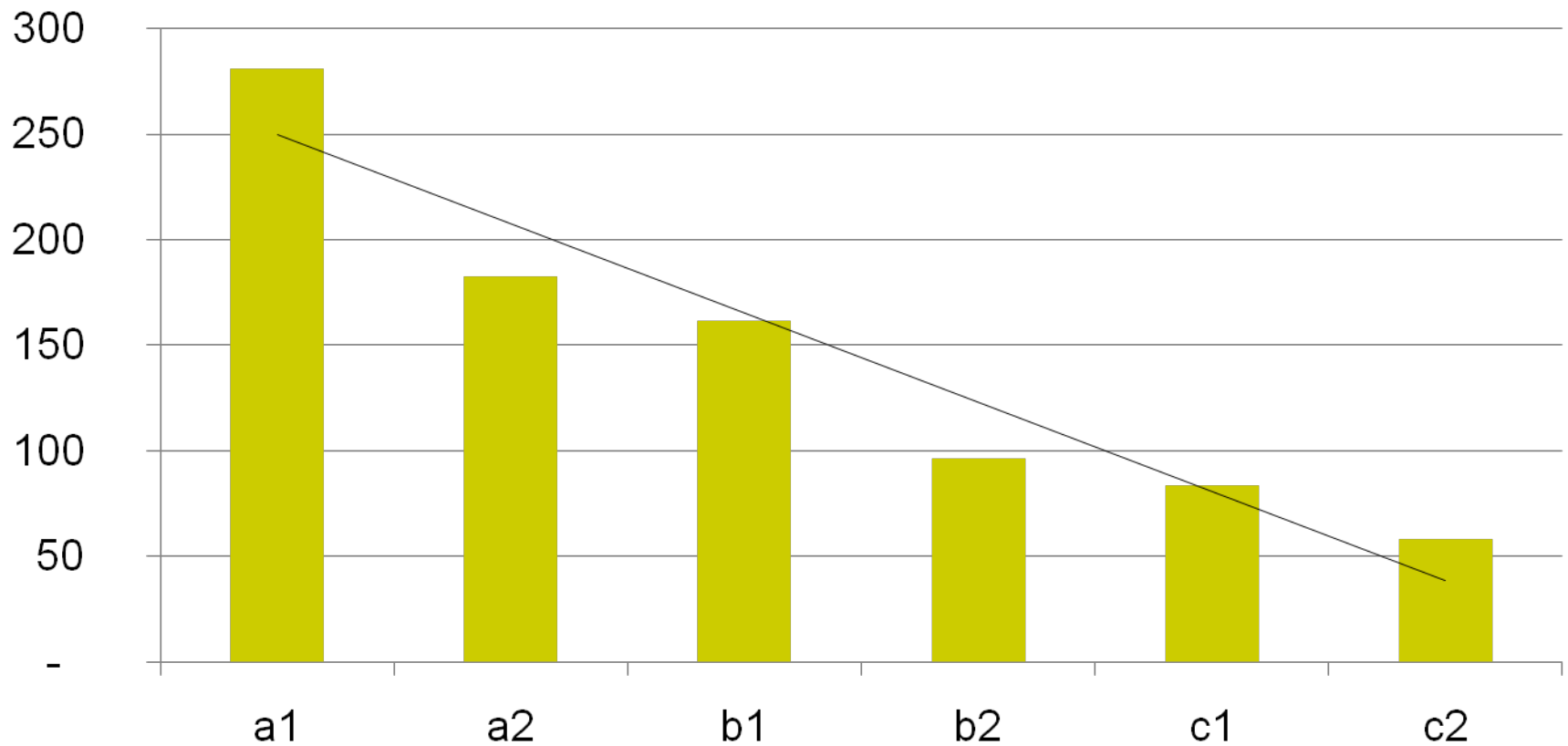


- Errors per level
- General Trend
- Main error categories
- Lexical errors per category

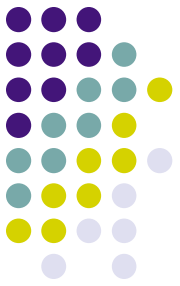
# Errors Made per Level



Errors / 1,000 words

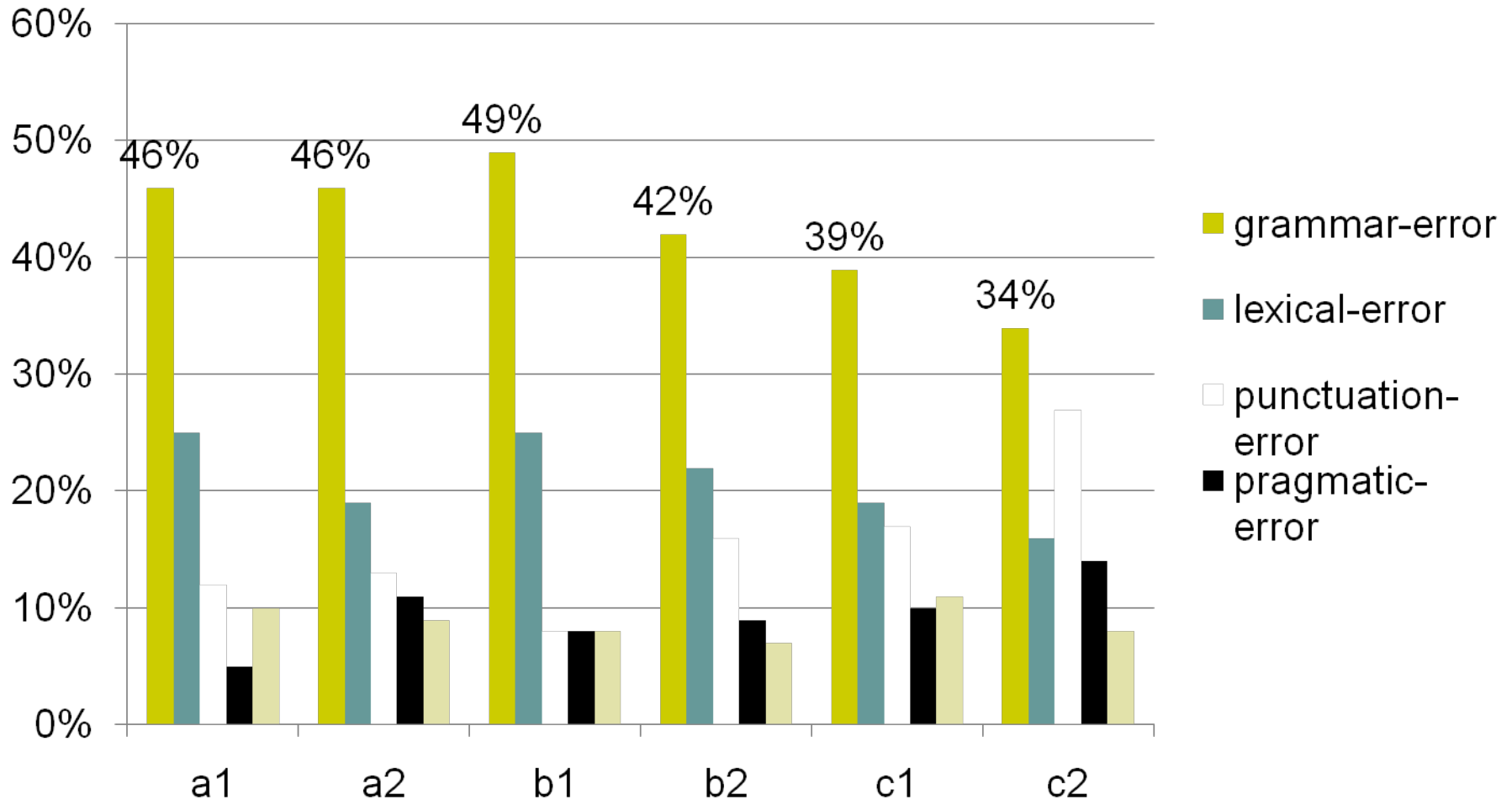
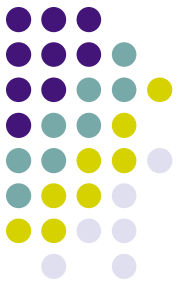


# Results per Level

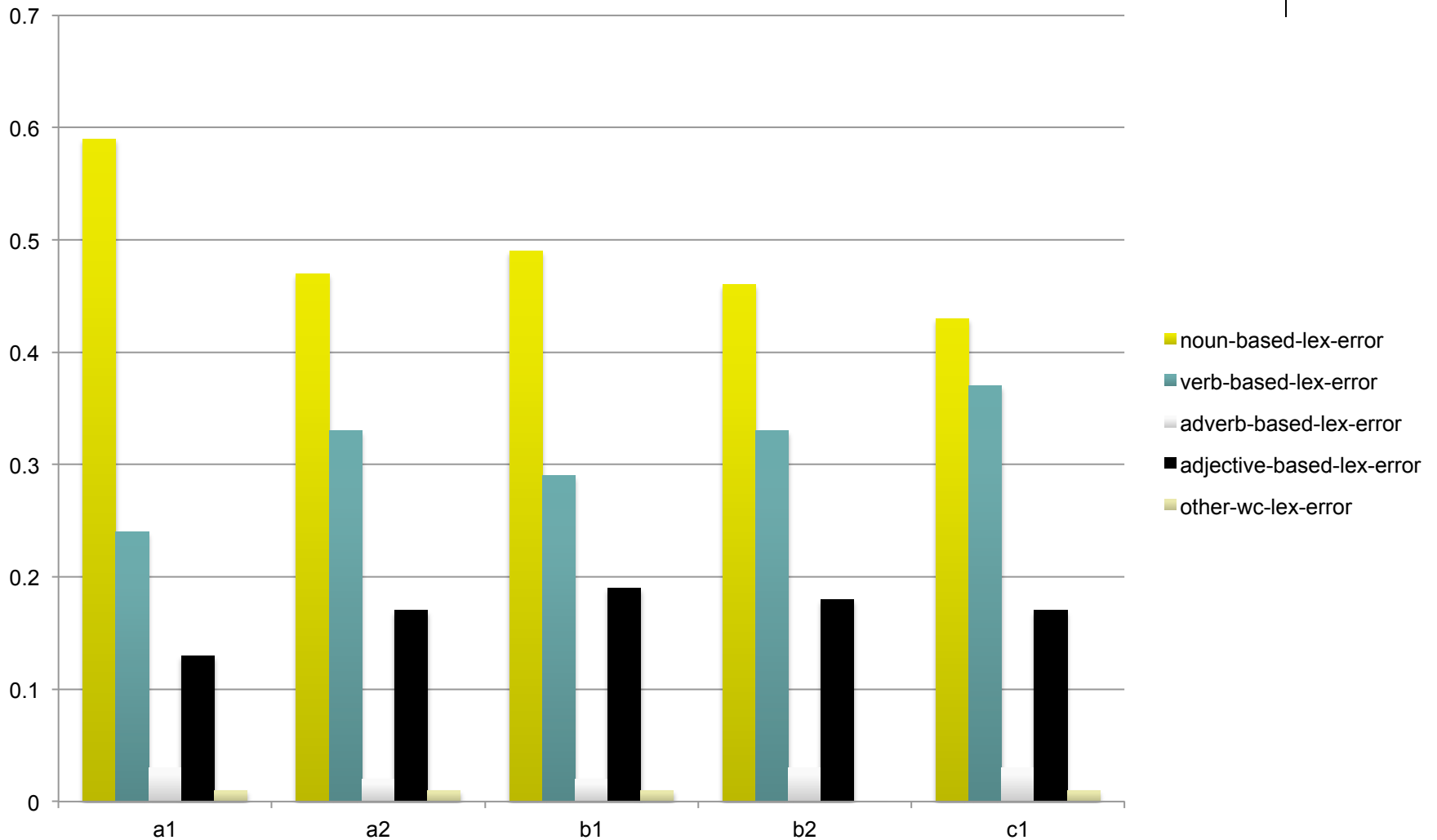
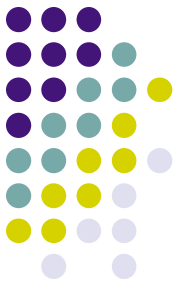


- Total Errors: 16,000 (110,000 words)
- Errors per CEFR Levels
  - A1 – More lexical errors
  - B1 – Nearly 50% grammatical errors.
  - C2 – Punctuation and pragmatic errors.

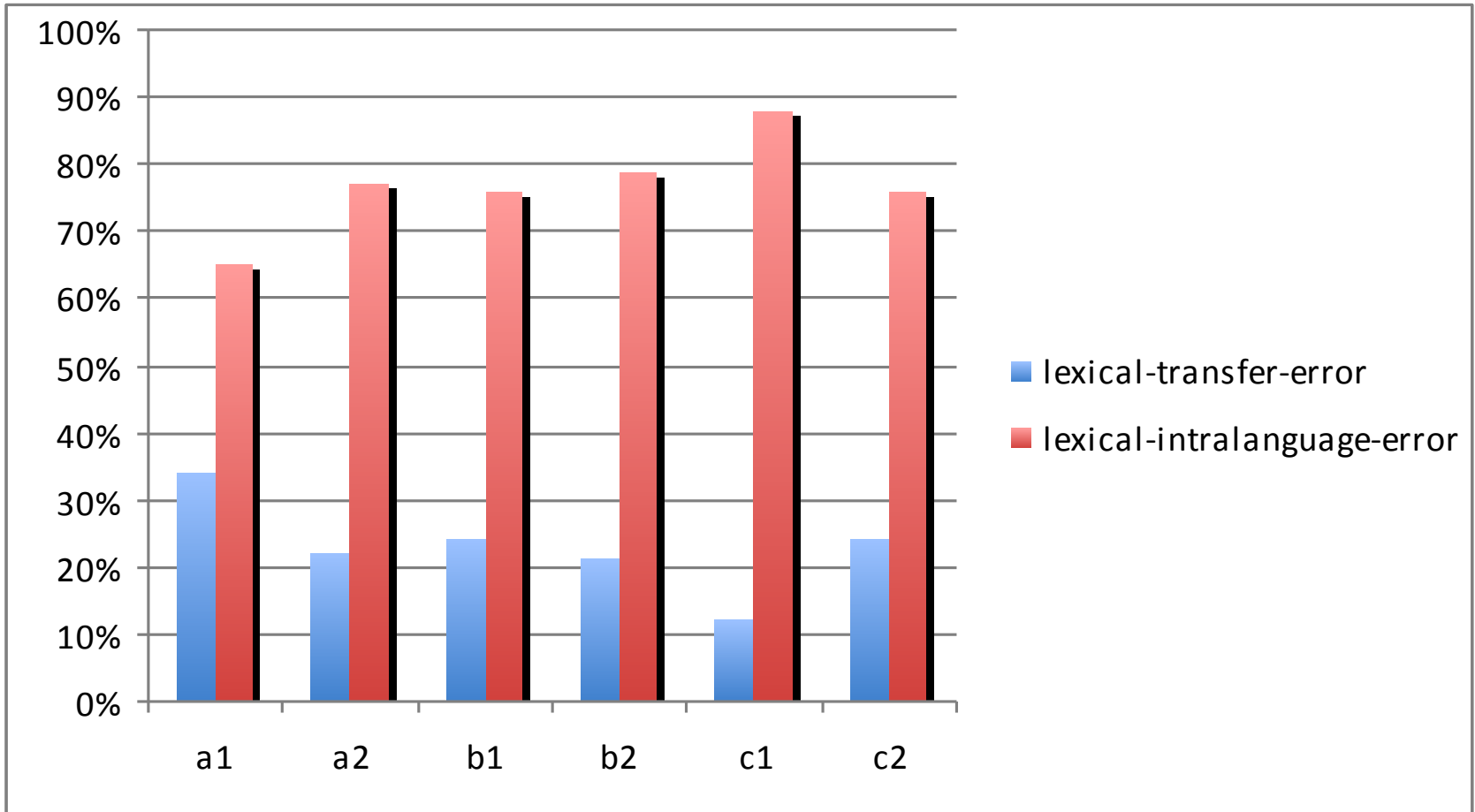
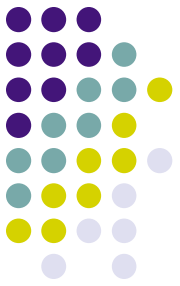
# Results per Level: Main Error Categories



# Types of Errors per Level



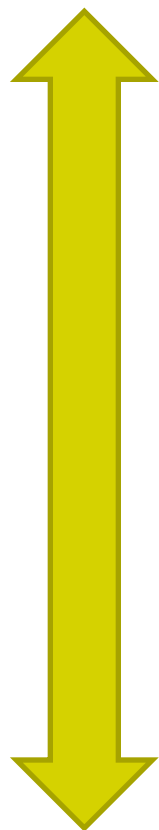
# Types of lexical errors per level



# Lexical Errors in terms of apparent difficulty



More common  
with basic  
learners



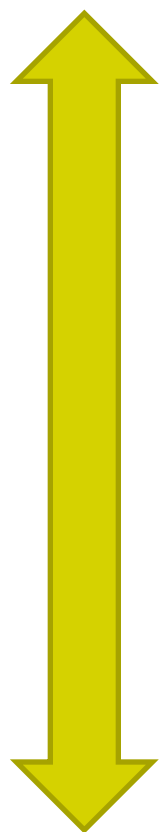
malformation  
coinage  
false-friend  
transferred-spelling  
verb-vocab-error  
spelling-error  
adverb-vocab-error  
borrowing  
noun-vocab-error  
adjective-vocab-error

More common  
with advanced  
learners

# Lexical Errors in terms of apparent difficulty



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

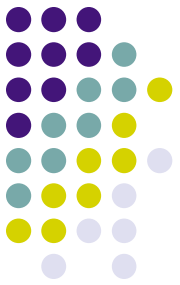
With the exception  
of borrowing,  
Transfer errors are  
more common for  
beginners, while  
later, intralanguage  
errors  
predominate.

More common  
with advanced  
learners

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

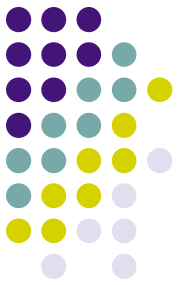


Transfer errors			Intralingual Errors	
Borrowing	Coinage	Transferred spelling	Spelling	Wordchoice
carril-bici laboral España ONGs Europa temporal mas hachis mundial conducta infantil habituate receptor	determined optative fomenting course (verb) sanity (health) poblation form displacements asignature desesperation diary principately evollution	inmigration inmigrant ilegal religión government possibilities cicles adiction tipes opinión politic costums asociation	live whit wich an (and) the a lifes countrys life foreing becouse there(their) beleive	persons work be other do make economical win have get job undeveloped doing



# 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** ([medieroesther@hotmail.com](mailto:medieroesther@hotmail.com))

**Web page TREACLE:** <http://www.uam.es/treacle>