

Error coding in the TREACLE project

Penny MacDonald, Susana Murcia, María Boquera, Ana Botella, Laura Cardona, Rebeca García, Esther Mediero, Michael O'Donnell, Ainhoa Robles, Keith Stuart



UNIVERSITAT
POLITÈCNICA
DE VALÈNCIA



Financiado por el Ministerio de Ciencia e Innovación (FFI2009-14436/FILO)



Introduction

Aim of paper:

- To present ongoing work on error coding in the TREACLE project



Error coding in the TREACLE project

- Introduction
- The TREACLE project
- Error coding process
- Intercoder reliability study
- Preliminary results
- Conclusions



What is the TREACLE project: Introduction

***TREACLE: Teaching Resource Extraction
from an Annotated Corpus of Learner English.***

- Participants in the project:
 - Universidad Autónoma de Madrid (UAM)
 - Universitat Politècnica de València (UPV)



What is the TREACLE project:

Aims of TREACLE:

- To profile grammatical skills
 - of Spanish university learners of English
 - at various proficiency levels
- To develop proposals for re-designing curriculum
- To develop teaching materials focused on Spanish students needs



What is the TREACLE project:

- Two stages in the first part of the project:
 - Automatic syntactic analysis
 - which structures the students are getting right in their output and at what level
 - which structures they are still developing; and which they do not yet use
 - Computer-aided error analysis
 - find out what they tend to get wrong at each level



What is the TREACLE project: Corpus

Two corpora:

- Wricle Corpus (UAM) (Rollinson & Mendikoetxea 2008)
- UPV Learner Corpus (Andreu et al., 2010)

Shared characteristics of corpora:

- metadata carefully recorded
- proficiency levels identified through Quick Oxford Placement Test (UCLES 2001)



The TREACLE project: Corpus

UPV Learner Corpus:

- Part of the MiLC corpus (Andreu et al., 2010)
- 950 written compositions (180,000 words)
- from UPV students of all levels
- Mostly on the topic of Immigration



The TREACLE project: methodology

Wricle Corpus (Rollinson & Mendikoetxea, 2008)

- 750 essays (approx. 1,000 words each)
- Spanish learners of English, from Filología Inglesa
- We use 521 of these essays (approx. 500,000 words)
- On topics of Immigration, Gay marriages and Traffic



The TREACLE project: methodology

- Error annotation using *UAM Corpus Tool* (O'Donnell, 2008)
 - Manual annotation, based on coding scheme
- Coding criteria document (20 pages long):
 - Criteria for segmentation
 - Criteria for annotation
- Inter-coder reliability studies

The TREACLE project: coding process

The screenshot shows a window titled "Error analysis for: Files/A101-4.txt". The main text area contains the following text:

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

Three callouts illustrate the coding process:

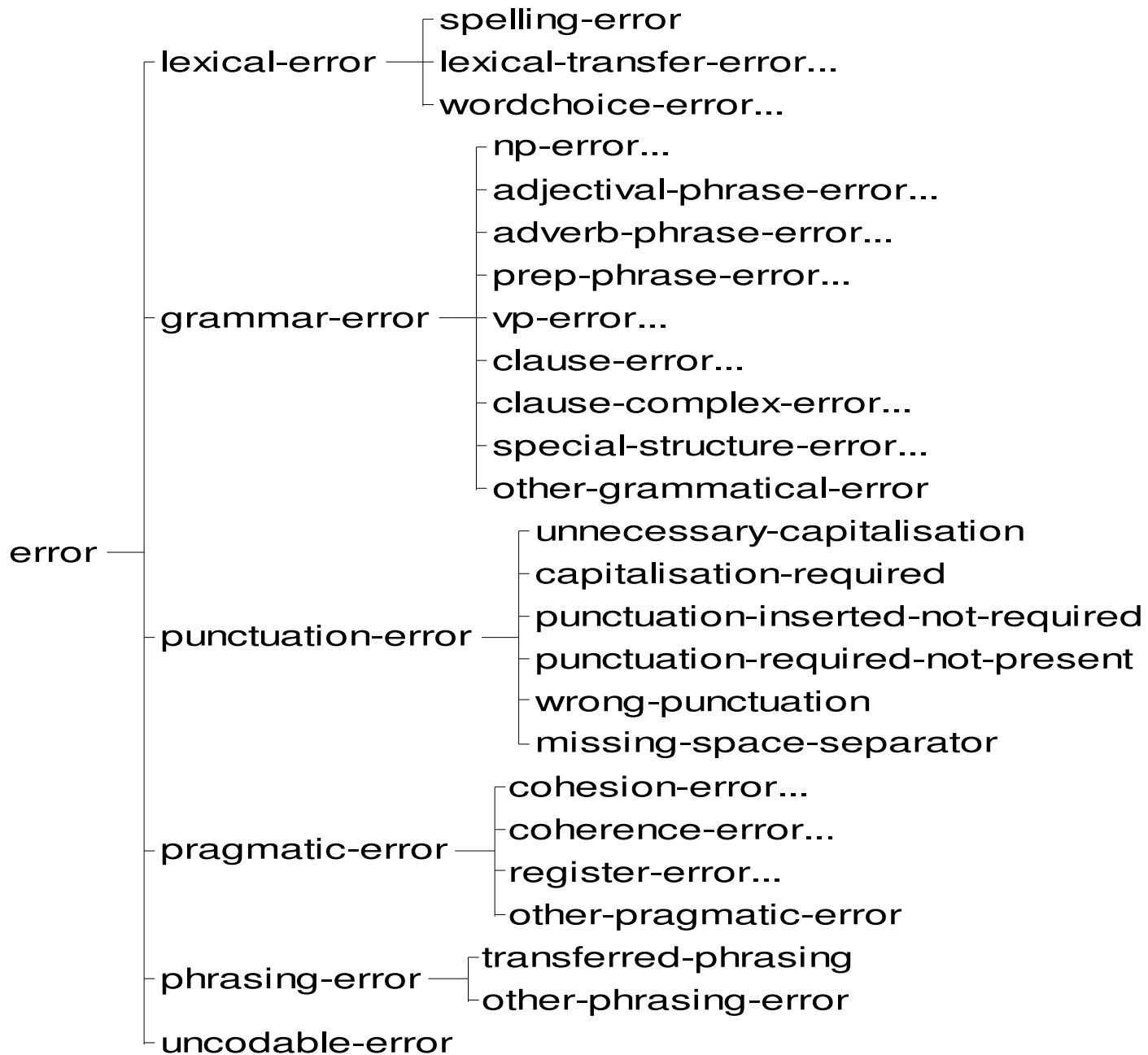
- 1. Select text containing error.** A callout points to the word "a" in the sentence "a whole society".
- 2. Provide the corrected text here.** A callout points to the "Correction:" field, which contains the text "the".
- 3. Assign features to current segment here.** A callout points to the "ERROR-TYPE" list, where "np-error" is selected.

The interface includes a toolbar with buttons for navigation (left and right arrows), "Ignore", "Delete", "Other Action...", "Save", "Close", and "Help". Below the toolbar are three columns: "Assigned" (containing "error" and "grammar-error"), "ERROR-TYPE" (containing a list of error types including "np-error", "adjectival-phrase-error", "adverb-phrase-error", "prep-phrase-error", "vp-error", "clause-error", and "clause-complex-error"), and "Gloss".



The TREACLE project: The error scheme

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



The TREACLE project: principles behind error scheme

- Error scheme maps onto organisation of grammar topics within EFL courses
 - Pedagogical goals – want to recover errors relevant to each topic
- Focused on the grammatical phrase where the error occurs, not on the part of speech of the error

e.g. *very browner

Adverb but functioning in an adjectival phrase

The TREACLE project: coding example

Example of error coding with *UAM CorpusTool*
*“this results are ...”

Major category type: **Grammatical** →
some grammatical rule is broken (wrong class
for slot, word order, agreement problem,
missing but necessary element, present but
unnecessary element, etc.)

Next level of delicacy:

Noun Phrase →

Determiner-error



The TREACLE project: how far are we in the coding process?

- Error coding is still at an early stage:
 - 126 texts coded, with 52,000 words
 - 6,540 errors
- Our aim is to double this by the end of 2011



The TREACLE project: intercoder reliability study

We performed two inter-coder reliability studies to:

- Refine the error scheme and coding criteria document.
- Ensure all coders were complying with the coding criteria.
- Average inter-coder reliability not high (~70%)
- But this is expected with:
 - Disagreements as to exact segment extent initially
 - Selection from 113 codes at most delicate level
 - In both studies, some participants had not been properly trained in using the scheme.

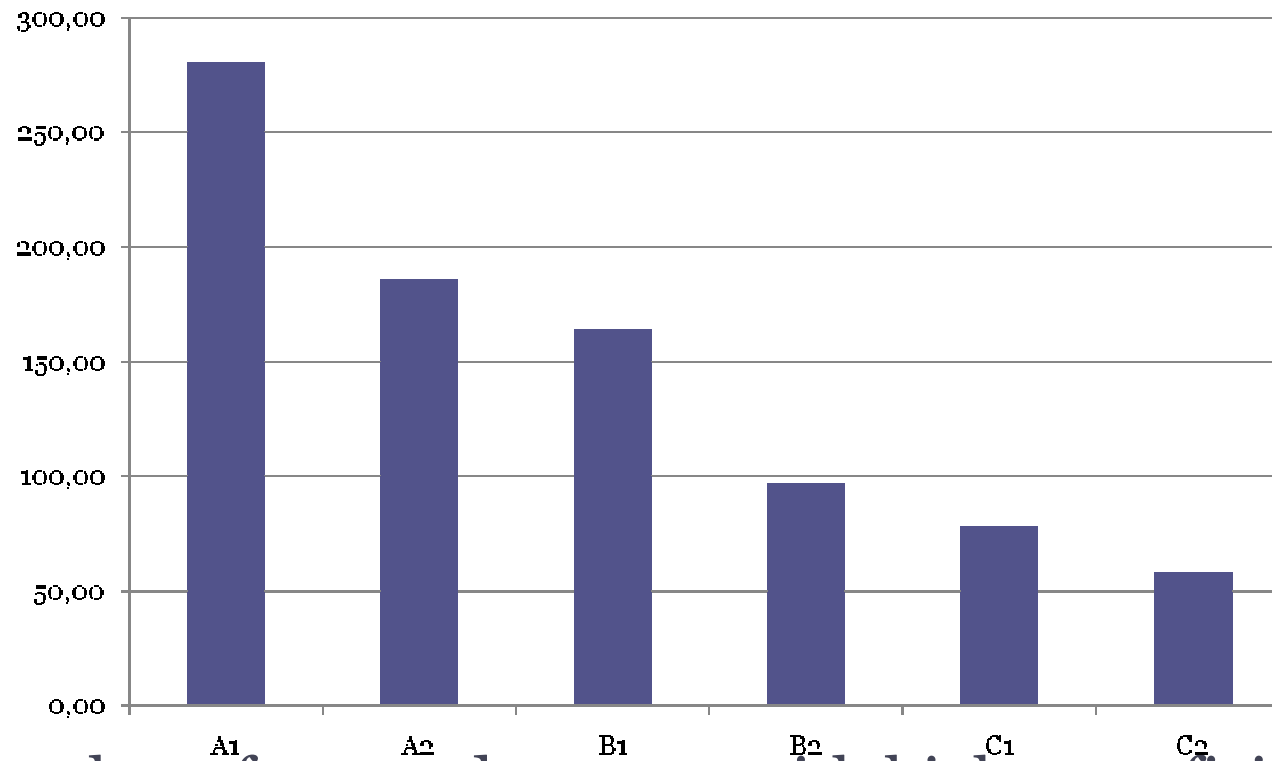


The TREACLE project: intercoder reliability study

Individual coding + consensus meetings

- Minimal segmentation, i.e. Select amount of text necessary to make correction of erroneous form:
*e.g. *in the other hand*
- Code as regards what the learner has written rather than what they should have written:
e.g. the car of John -> John's car: determiner problem or postmodifier problem? - principle of coding what students do - postmodifier error.

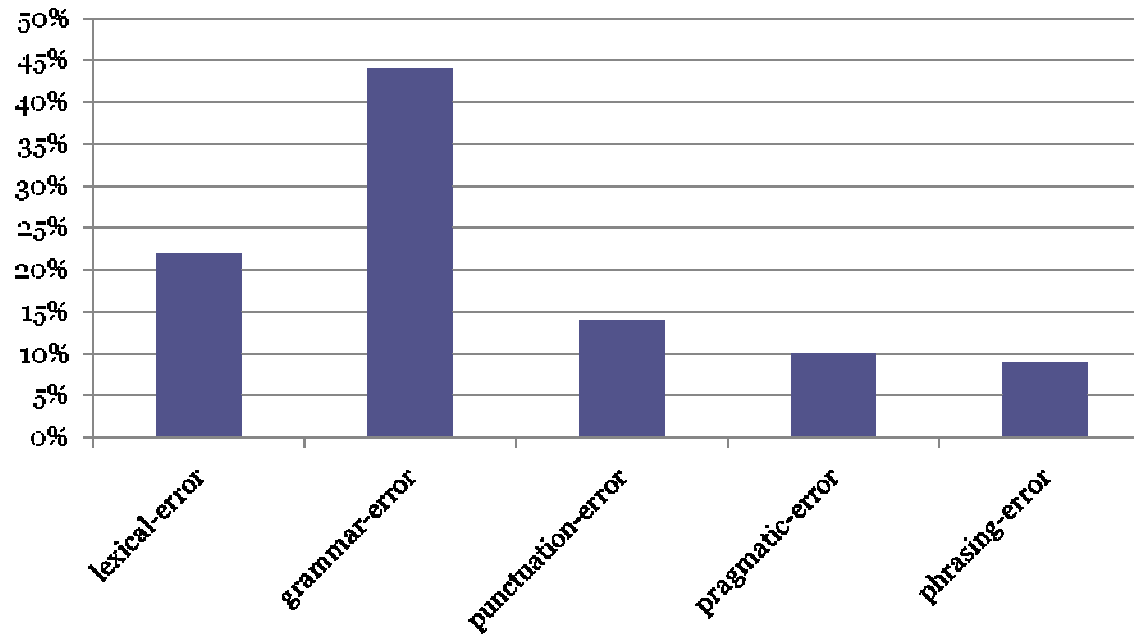
The TREACLE project: some preliminary results of coding process



Number of errors decreases with higher proficiency level

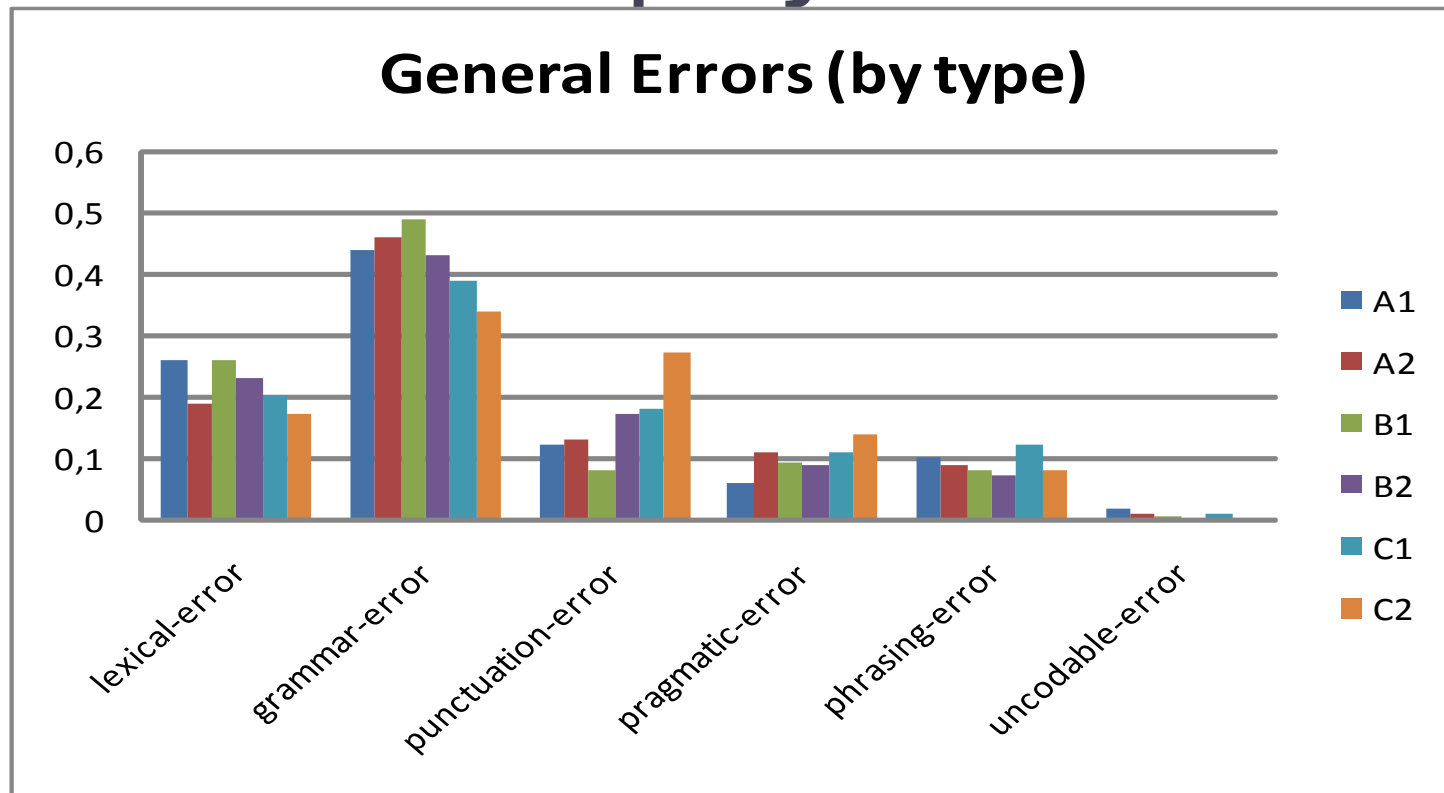
The TREACLE project: some preliminary results of coding process

Errors in General



Grammar errors account for 44% of all errors

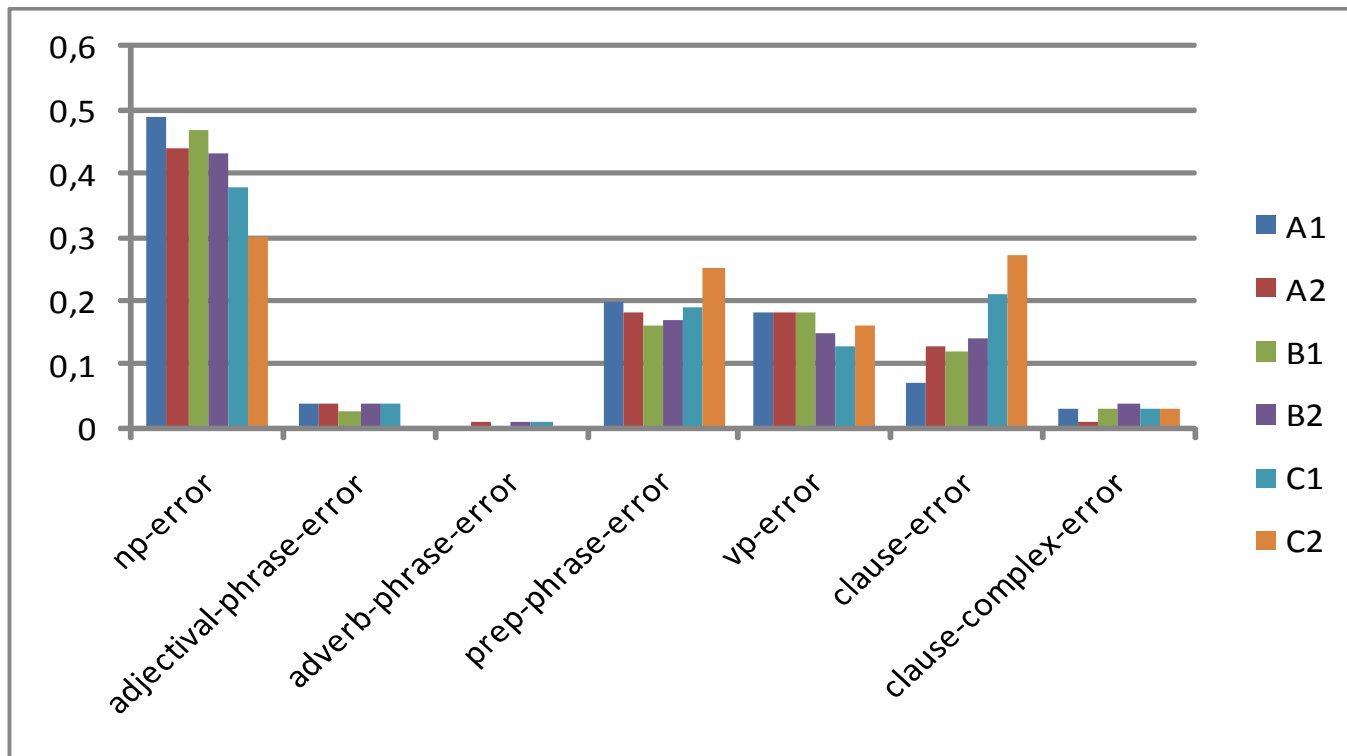
The TREACLE project: results



- A1s and B1s highest number lexical errors, which then decline as proficiency increases.
- Grammar errors increase steadily from A1 to B1, then drop.

The TREACLE project: results

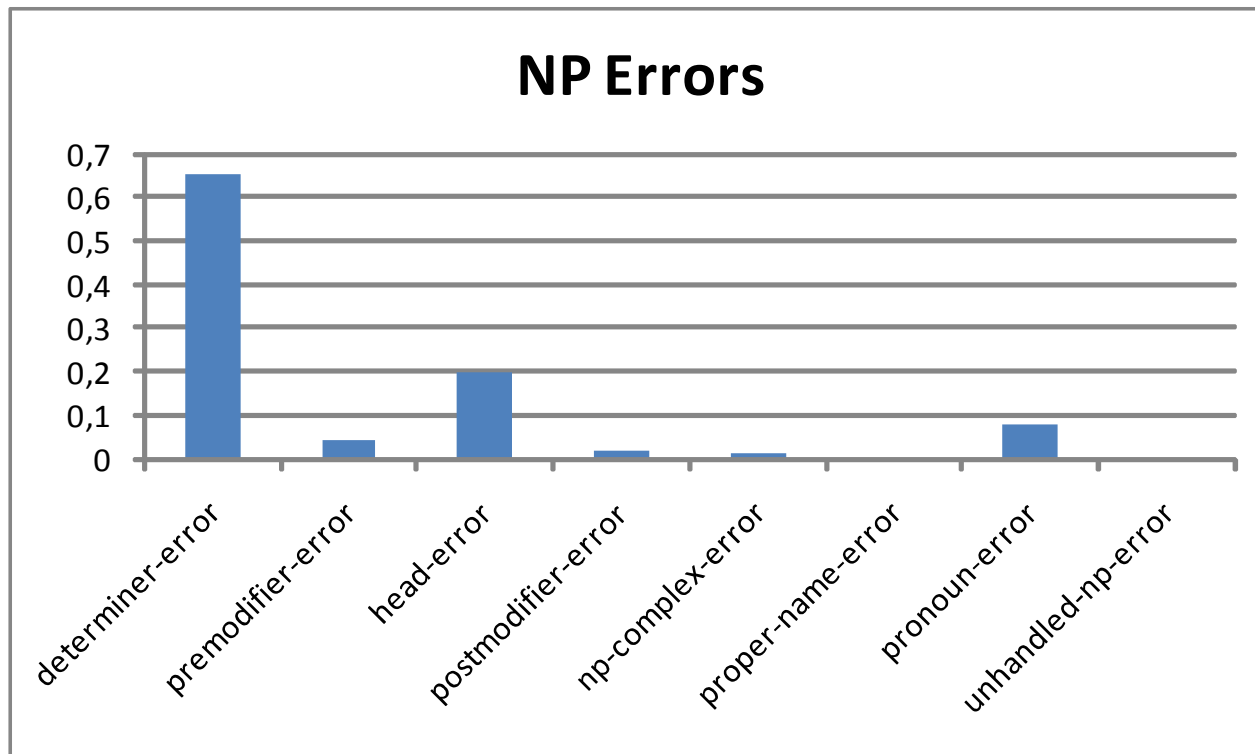
Grammar errors at each level



- NP errors high – improves with proficiency
- PP next highest within Grammar category

The TREACLE project: results

NP errors



Determiners = 65% of NP errors and 30% of all grammar errors. How much time is spent on these in EFL programmes?

The TREACLE project: results NP errors - Determiner search

Enter Search Query Below:

clauses + containing immediately noun-phrase + containing immediately determiner-present-not-required +

P Only Partially Coded ***** Segment with Comment 29 matches

	P	C	Pretext	Contained	
	F	*	o contaminating movements, increase in	a 30 %	the rail bus, develop even more
	F	*	a 30 % the rail bus, develop even more	the public transport	
	F	*	Secondly,	the education at school	must be considered as another
	F	*	ces, they have reached their position on	the society	, through a lot of demonstration
	F	*	having the same rights like	the other people	
	F	*	to adopt	some children	
	F	*	that	the children	go to be adopted by the couple
	F	*	e twenty of december, Holand legalized	the marriage	and the adoption.
	F	*	ber, Holand legalized the marriage and	the adoption	.
	F	*	and let too	the marriage and adoption	.
	F	*	In Canada in	courious, because the civil marri	but the church marriage is com
	F	*	that the amount of cars within	the population	is increasing



The TREACLE project: What can we do with these results?

- Use these profiles to adjust the teaching curriculum: determining which grammatical features need to be taught, in what order, and with what degree of emphasis.



The TREACLE project: Conclusions (1)

- Methodology
- Results obtained from the error coding
 - Number of errors decreases with greater proficiency
 - Types of errors by students change at different levels of proficiency



The TREACLE project: Conclusions (2)

- Potential applications of the study –
- Typical amount of teaching given to grammar topics - not always proportional to the learners' needs as shown by the errors they make:
 - Large proportion of determiner errors, but determiners not given much attention
 - Minimal teaching of prepositions, but students struggle with them at all levels
 - Teaching of verb phrase given too much time?



The TREACLE project: Suggestions and further research

- We propose to reinforce those grammar areas where errors are made
- Need to distinguish between explicit teaching of concepts vs. on-line drilling and practice
- Error analysis not in isolation – it needs to be seen in the context of what the students are attempting and getting right.
- Later work will explore learner types based on the experimental vs. cautious learners.



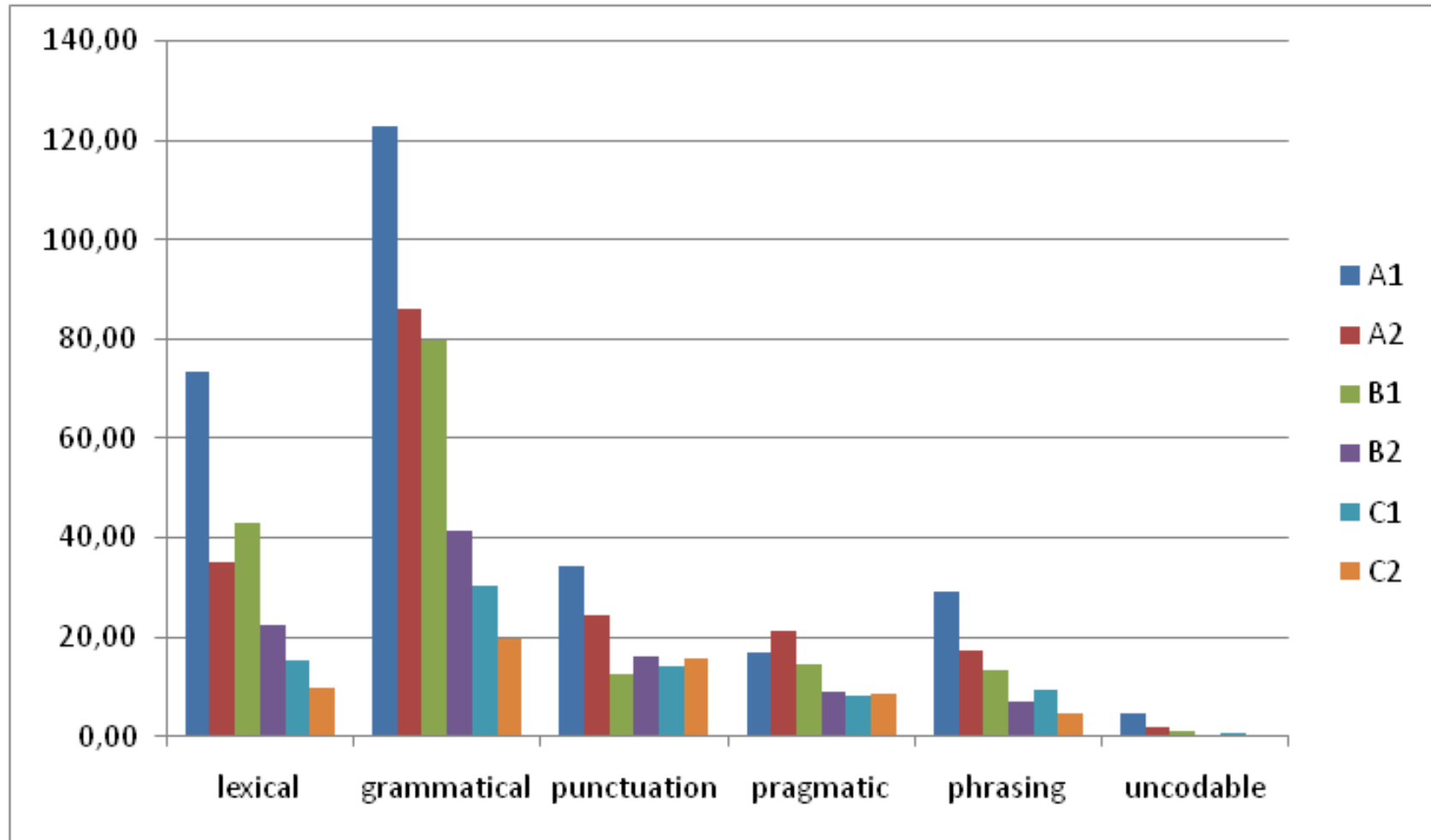
Thank you

Penny MacDonald (penny@idm.upv.es)

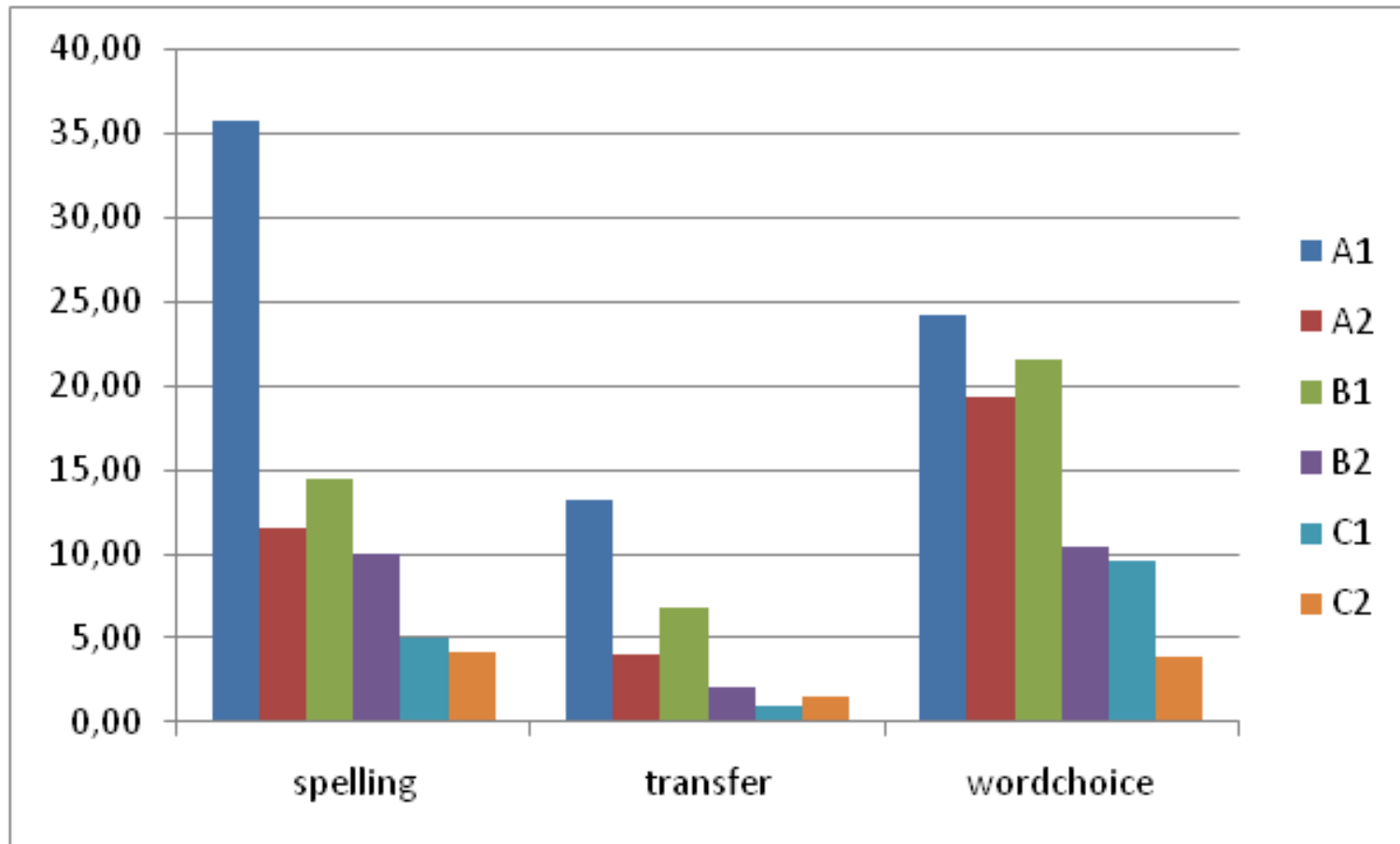
Susana Murcia (susana.murcia@uam.es)

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

Number of general errors per 1,000



Subtypes of lexical errors per 1,000 words





Errors coded by researchers in TREACLE to date:

- ✓ 126 texts (total words: 55,583)
- ✓ 17 by A1 (2,436 words)
- ✓ 39 by A2 (11,392 Words)
- ✓ 26 by B1 (9,713 Words)
- ✓ 25 by B2 (13,620 Words)
- ✓ 14 by C1 (13,322 Words)
- ✓ 5 by C2 (5,100 Words)

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

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**’.*

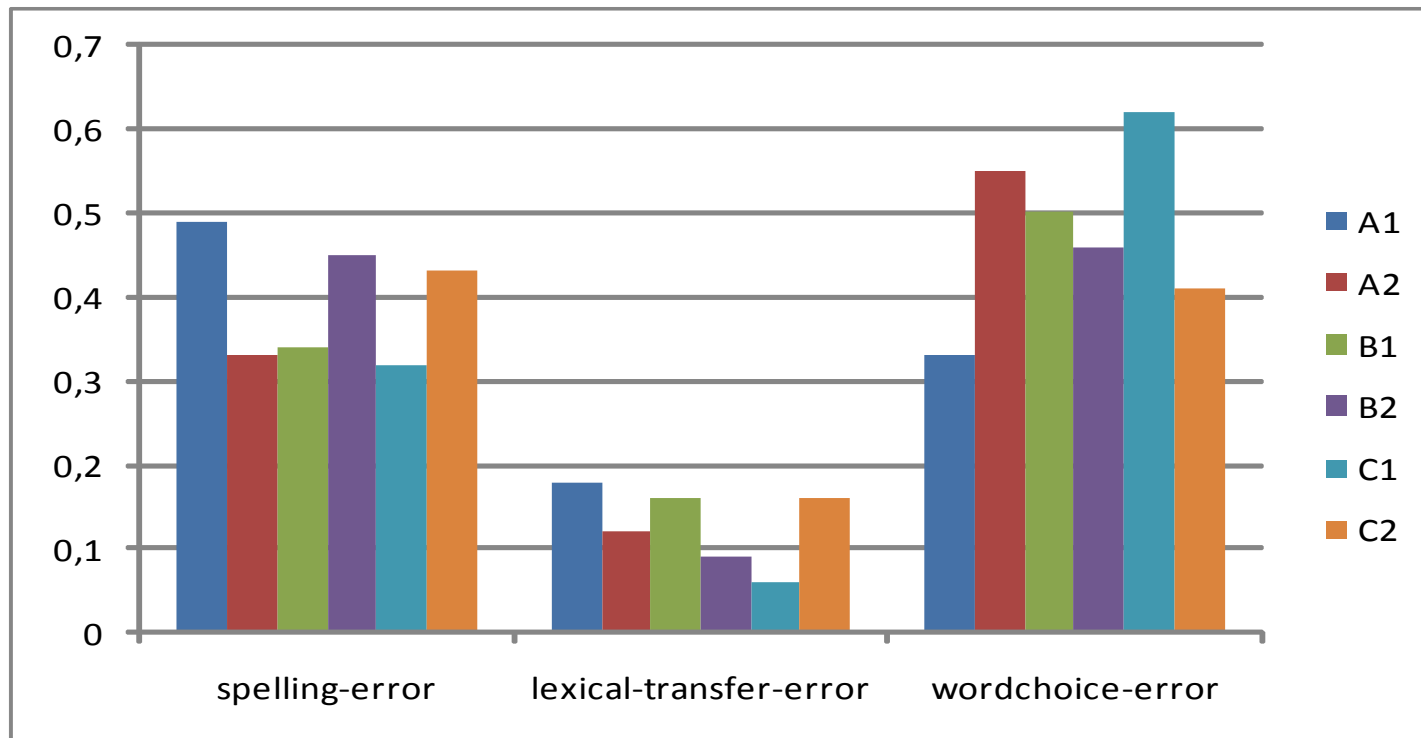
- Disagreements like this were solved during the consensus meetings

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	✓

The TREACLE project: results

Lexical errors



- Errors in spelling and lexical transfer – A1 .
- Word choice, generally decreases with proficiency.
- Note: spelling errors can be typos (carelessness) or lack of knowledge of TL forms.