



TREACLE:

Improving the grammatical competence of Spanish university EFL students: design and implementation of a web-based learning system.

Mick O'Donnell

Penny MacDonald

**WORK IN
PROGRESS**



The TREACLE Project

Project: TREACLE



Teaching
Resource
Extraction from an
Annnotated
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 2010-2013
(FFI2009-14436/FILO)

The TREACLE Project

What are the general goals of the TREACLE blended learning Project?

- Detailed study of vocabulary and grammatical skills of students studying English at Spanish universities.
- AIMS at efficient and quality teaching tailored to meet students' individual needs.
- Focus on reform of grammar teaching both in the traditional classroom and also in online learning.

Towards Targeted Learning I

The "shotgun" approach to learning

- Many CALL systems take a **shotgun** approach to learning:
 - They have a general idea where the user is,
 - They teach language concepts (grammar, vocab, discourse, etc.) over that area.



Towards Targeted Learning I

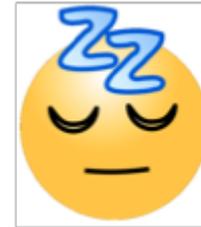
Problem with the "shotgun" approach:

- Each learner is an individual - they have different learning experiences, learning aptitudes, etc.

- Some of the language material will cover concepts they already know



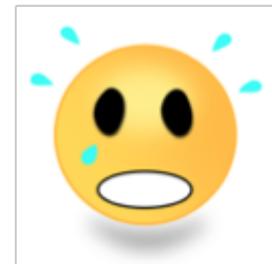
BOREDOM



- Some of the material will cover concepts they are not yet ready for



CONFUSION



Towards Targeted Learning I

A Targeted Learning Approach

Solution: **targeted learning:** the online system should teach exactly those language concepts that are most critical to the current learner at this point of time.

Learners learn best when they see a clear relevance to their needs.

Towards Targeted Learning I

A Targeted Learning Approach

Critical concepts: two aspects of criticality

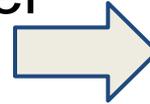
1. Language concepts which peers of the learner know but the learner does not.
2. Language concepts that are real and observable problems to language learners.

Towards Targeted Learning I

A Targeted Learning Approach

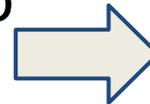
Critical concepts: two aspects of criticality

1. Language concepts which peers of the learner know but the learner does not.



Study of learner texts ranked over proficiency levels to see where concepts typically are an issue

2. Language concepts that are real and observable problems to language learners.



Study of learner texts to see which language features most frequently lead to errors.



Discovering Critical Grammar Concepts for Spanish Learners of English

Towards Targeted Learning I

A Targeted Learning Approach

Importance of using criticality in language teaching:

- Learning a foreign language requires mastering thousands of grammatical concepts.
- But many of these concepts are easily acquired, directly transferable from the mother tongue, e.g., English and Spanish share passive structure, progressive aspect, etc.
- So, observing where they go wrong (errors) can help us identify what parts of the language are critical for their learning.
- Also, comparing frequency of use of grammatical structures in learners against comparable natives show how much effort is needed for each problem.

Methodology - the corpora

The project involves 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 Andrés et al. 2010)

Oxford Placement test given at same time, to measure proficiency

Methodology

A two-pronged approach was followed:

- Manual analysis of learners' **errors**

Coded: 307 essays, 113,000 words, **16,200** errors

 What learners do wrong.

- Automatic analysis identifying **syntactic** structures used by the learners:

Coded: 1330 essays, 700,000 words, **98,000** clauses

 What learners are doing / not doing

Manual and automatic parsing done via UAM CorpusTool,
developed by the project and available from:

<http://www.wagsoft.com/CorpusTool/>

Error Annotation

- Errors tell us what problems learners face in their writing (and thus that need to be addressed by an online system)
- We explore errors in relation to **proficiency** levels to assign an "order of difficulty" to the language feature behind the error.
- We explore errors in terms of **frequency** to determine how much emphasis the online system should give to each feature.

The Error-coding Process

The screenshot shows a software window titled "Error analysis for: Files/A101-4.txt". The main text area contains two paragraphs. The first paragraph has a small green cursor under the word "a" in "a whole society". A callout box labeled "1. Select text containing error" points to this word. Below the text is a toolbar with buttons: <<, <, >, >>, Ignore, Delete, Other Action..., Save, Close, and Help.

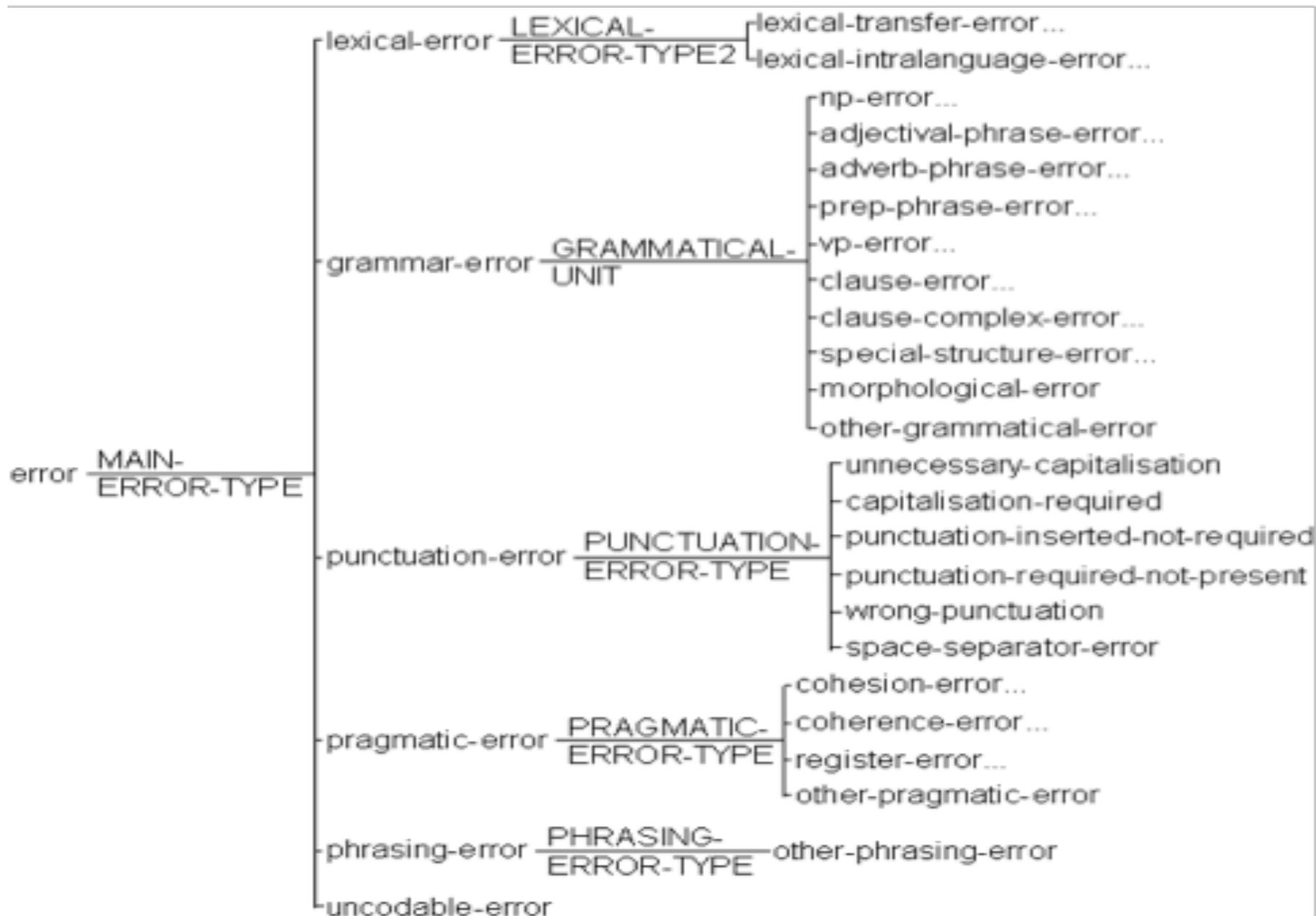
Below the toolbar are three columns: "Assigned", "ERROR-TYPE", and "Gloss".

- The "Assigned" column has a list with "error" and "grammar-error".
- The "ERROR-TYPE" column has a list with "np-error" (highlighted), "adjectival-phrase-error", "adverb-phrase-error", "prep-phrase-error", "vp-error", "clause-error", and "clause-complex-error". A callout box labeled "3. Assign features to current segment here" points to this list.
- The "Gloss" column is empty.

At the bottom, there are two input fields:

- "Correction:" with the text "the" entered. A callout box labeled "2. Provide the corrected text here" points to this field.
- "Comment:" with an empty text box.

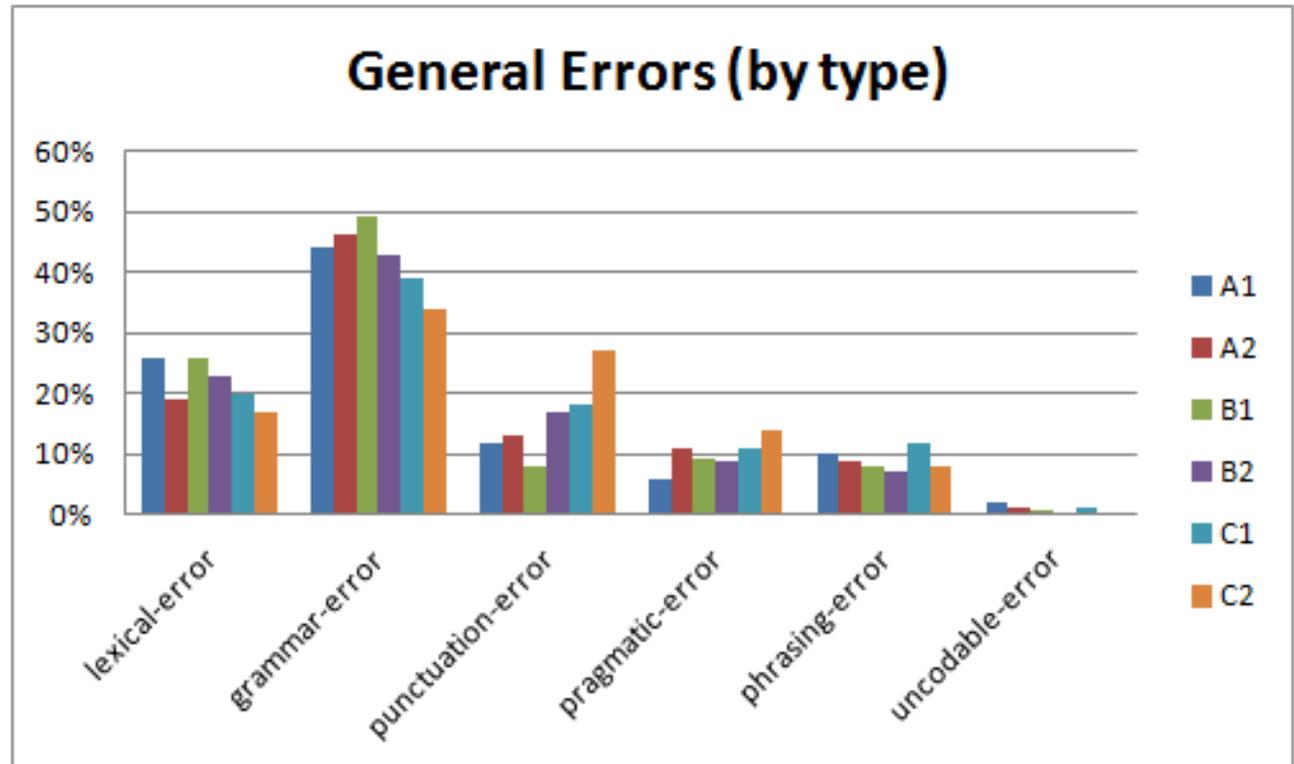
Error Annotation



Results of error analysis

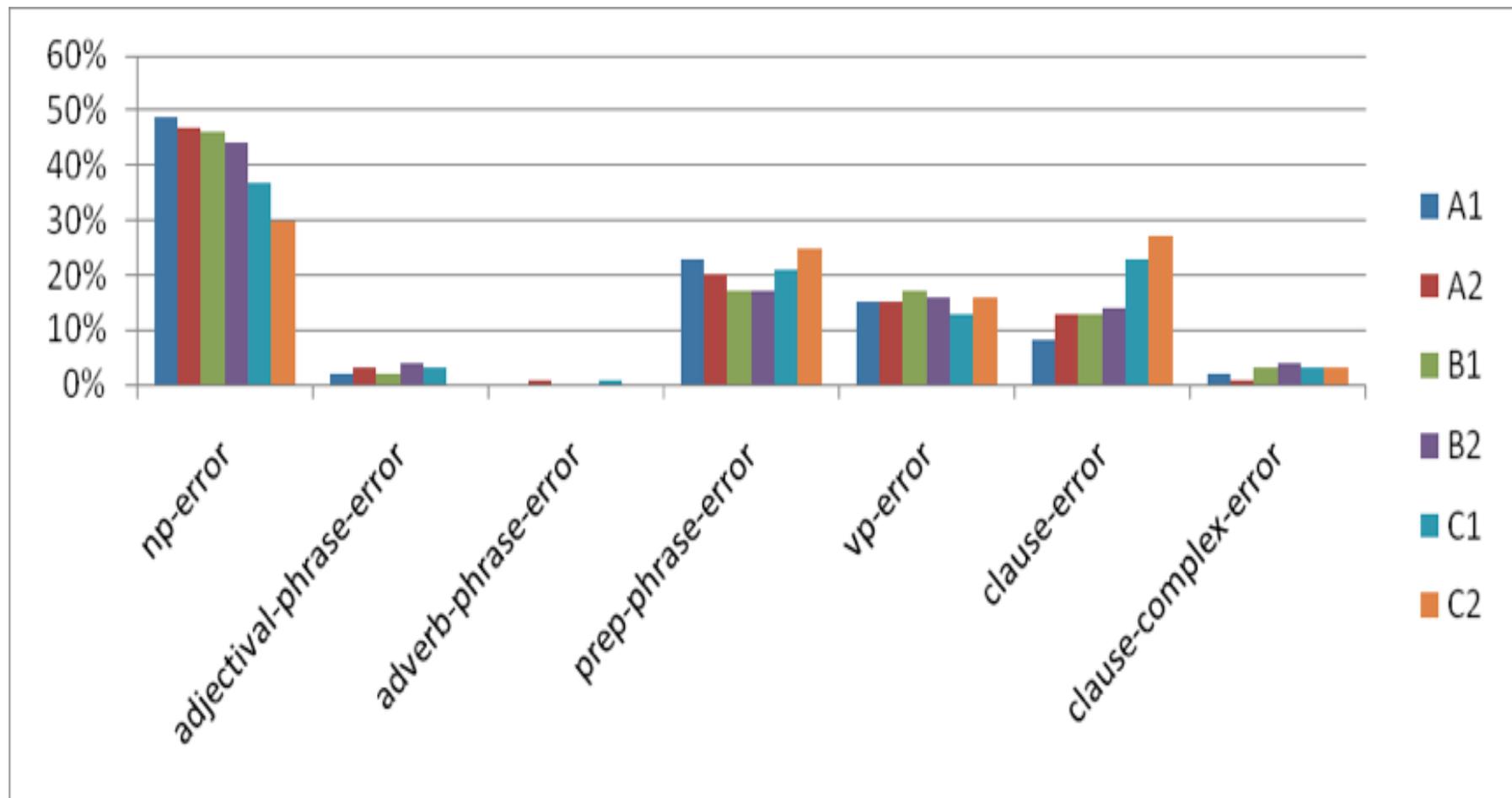
Errors per type

Grammar errors account for nearly 50% of all errors at B1 level, which is reduced to 34% of total at C2 level.

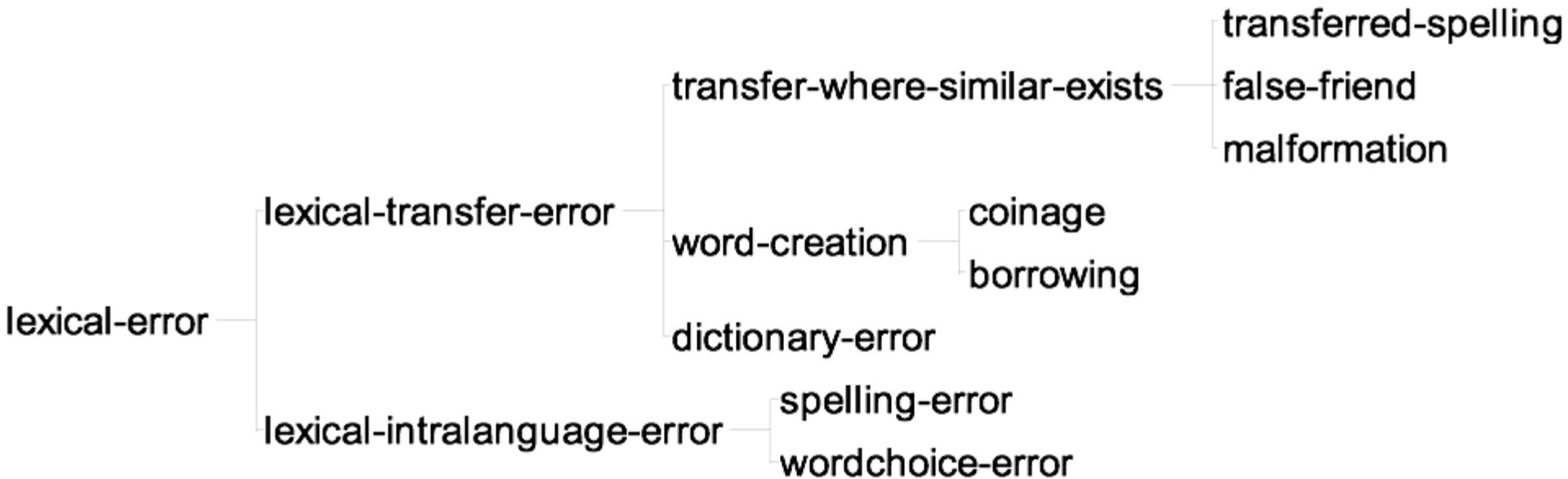


Results of error analysis

Grammar errors - different sub-categories



Error Annotation - Lexical errors



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.

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

More common with advanced learners

Error Annotation

Grammatical Errors in order of Frequency:

- determiner-present-not-required 1087
- prep-choice-error 818
- subject-finite-agreement 535
- determiner-absent-required 438
- wrong-number 428
- determiner-choice-error 248
- determiner-agreement 230
- obligatory-subject-absent 226
- unnecessary-preposition 204
- adjunct-order-error 177
- pronoun-choice-error 134
- ...

Syntactic Annotation

- Corpus has been automatically structurally parsed by Stanford parser
 - *1330 texts, 700,000 words, 98,000 clauses,*
- Syntactic features extracted by pattern recognition (e.g. active/passive, ditransitive, past-perfect etc.)
- We measure the degree of occurrence of each grammatical feature in relation to rising proficiency score to order grammatical features in terms of difficulty. (See O'Donnell 2013)
- **Work to come:** extending the range of grammatical structures recognised to include a wider range of critical concepts.

Automatic Syntactic Annotation

Grammar analysis for: Files/A101-2.txt

The new points system for driving offences will be established in Spain before summer o...

Subject				Mod	Pass	Pred	Adjunct			
Deict	Epith	Thing	Thing	Qualif			Op	Pphead	Op	
			Op	Pphead			Thing			Thing
			Classif			Thing				

With this new system , the driving licence will consist of a number of points that c...

Adjunct			Sep	Subject			Mod	Pred	Adjunct			
Op	Pphead			Deict	Classif	Thing			Op	Pphead		Conj
Deict		Epith	Thing						Deict	Thing	Qualif	
									Op	Pphead		
									Thing			

I personally agree with the establishment of this new law , as I feel tha...

Subject	Adjunct	Pred	Adjunct					Sep				
Thing	Head		Op	Pphead				Conj				
			Deict	Thing	Qualif				Subject	Pred		
					Op	Pphead			Thing			
					Deict	Epith	Thing	Cor				

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

Assigned

- grammatical-unit
- group
- np
- common-phrase
- singular-phrase
- nonwh-noun-phrase

Gloss

Comment:

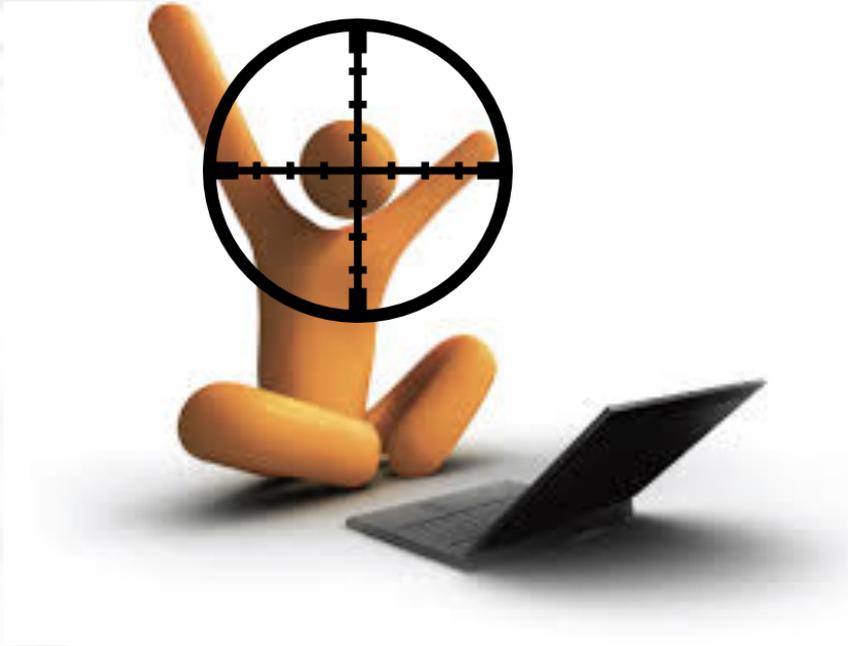
Annotation: Summary

So, We have:

- 16,000 manually annotated errors
- 98,000 clauses syntactically tagged

From this, we discover:

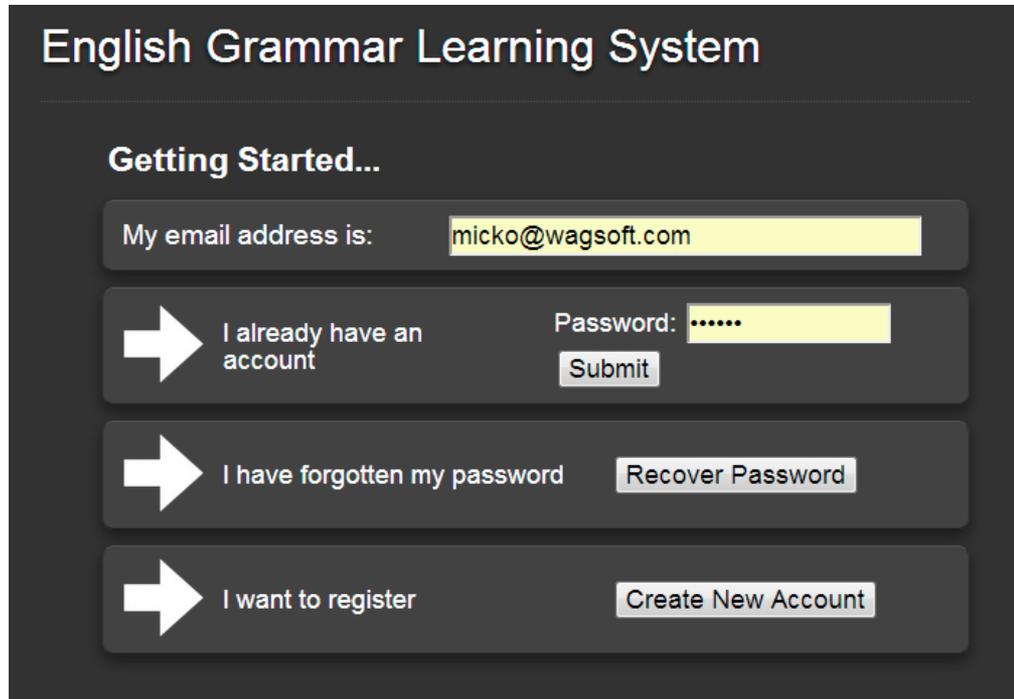
- **Frequency of occurrence** of grammatical errors and thus importance of the connected concepts
- **Order of difficulty** of grammatical concepts, both from when learners are making errors in them, and how often they produce them at each proficiency level.



Towards Targeted Teaching

Towards Targeted Learning II

- A "test of concept" web-system has been produced.
- Plan to deliver to our first year students for 2013-14
- Organisation of grammatical topics in terms of a topic hierarchy



The image shows a dark-themed user interface for an "English Grammar Learning System". The title "English Grammar Learning System" is at the top. Below it is a section titled "Getting Started...". There are three main input areas, each with a white arrow icon on the left. The first area is for email, with the label "My email address is:" and the value "micko@wagsoft.com" in a yellow box. The second area is for login, with the label "I already have an account", a "Password:" label, a yellow box with ".....", and a "Submit" button. The third area is for password recovery, with the label "I have forgotten my password" and a "Recover Password" button. The fourth area is for registration, with the label "I want to register" and a "Create New Account" button.

English Grammar Learning System

Getting Started...

My email address is:

 I already have an account Password:

 I have forgotten my password

 I want to register

Towards Targeted Learning II

Modelling the Language...

- Language Model: For each grammatical concept:
 - The order of difficulty of the concept
 - The frequency of occurrence of errors related to the concept

(only partially implemented)

Concept	Gloss	Example	Counter-example
much-not-in-simple-decl	'much' is not usually used in affirmative sentences.		He has much water.
much-with-mass	'much' is used with mass nouns only.	much water	much apples
much-with-neg	'much' can be used in negated statements.	I don't have much money.	
much-with-neg-embedded	'much' can be used in a positive clause embedded in a negative one.	I don't think he has much money.	
much-with-question	'much' can be used in questions.	Do you have much money?	
much-with-intens	'much' can be used in affirmative sentences when it is preceded by "so", "too" or "as" (intensifiers).	I have too much water.	

The Question Database

- a database of multiple-choice type questions.
- For each answer, indication of the grammatical concepts confirmed or broken.

[] his legs were injured in the explosion.

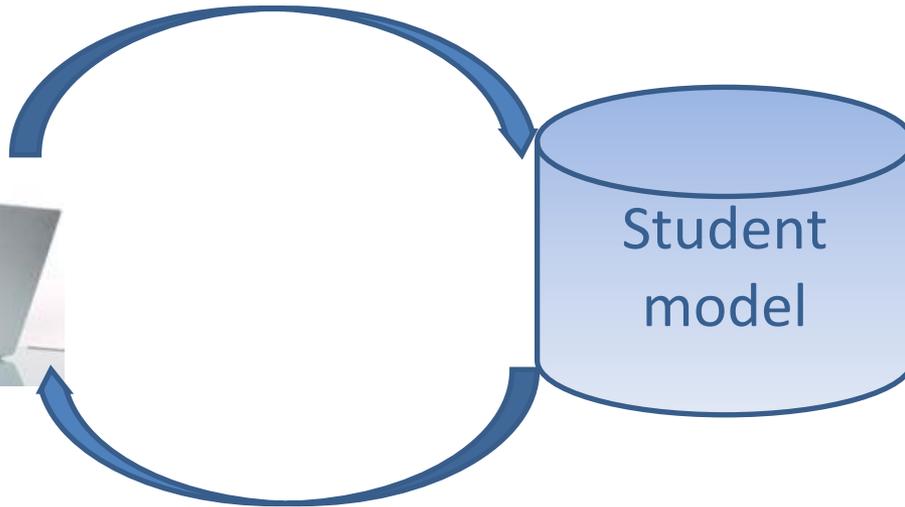
Answer	Concepts Broken	Concepts Complied
<i>neither</i>	neither-cant-be-nonpartitive-predet	use-dual-form-where-possible; neither-is-dual
<i>the two</i>	np-cant-be-nonpartitive-predet; use-dual-form-where-possible	
<i>both</i>		both-can-be-nonpartitive-predet; both-is-plural; both-is-dual use-dual-form-where-possible;
<i>either</i>	either-is-singular; either-cant-be-nonpartitive-predet	use-dual-form-where-possible; either-is-dual

Towards Targeted Learning II

Modelling the LEARNER...

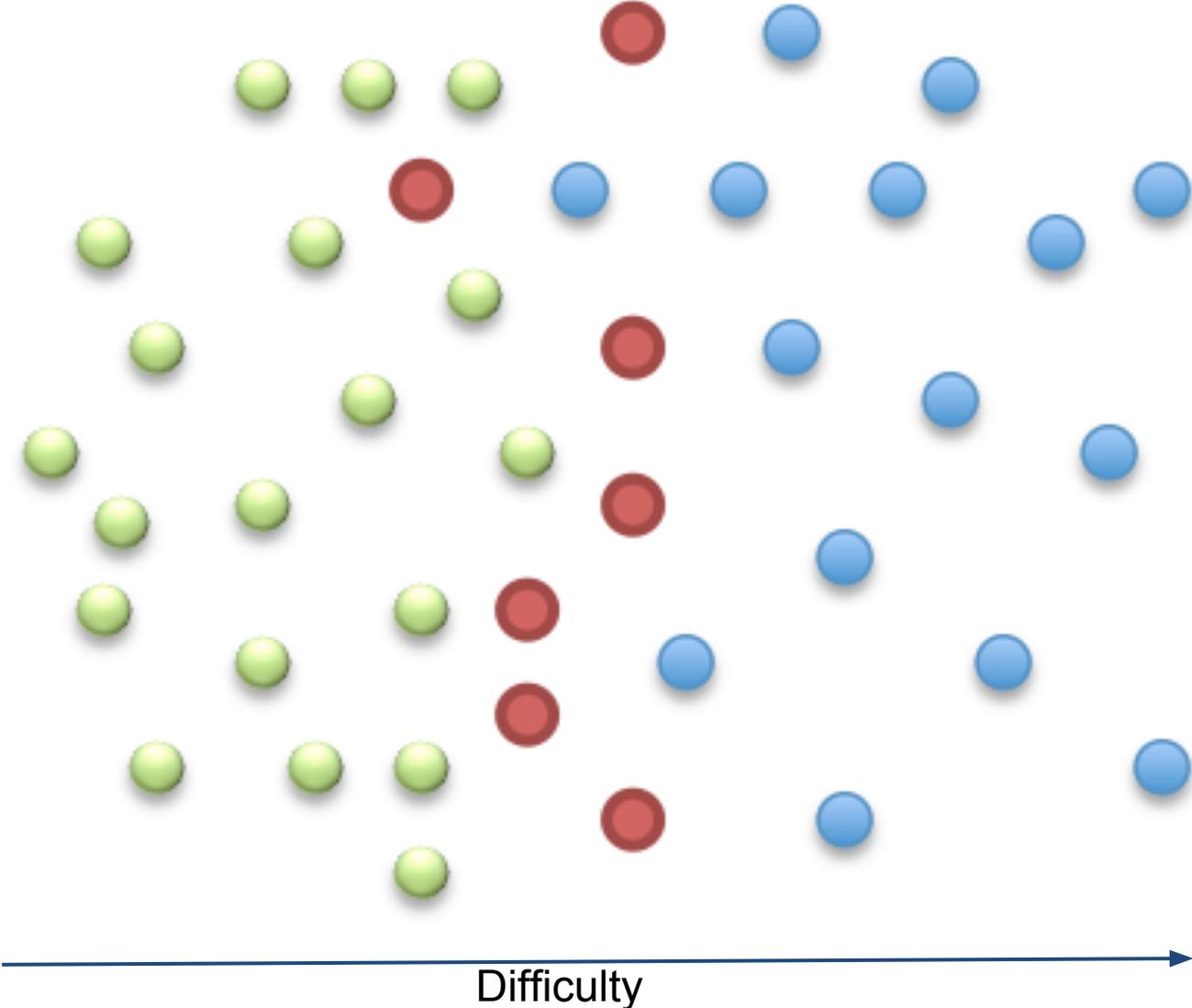
- Learner Model: showing, for each learner:
 - the level of assimilation for each grammatical concept
 - the response history for each quiz question

Student responses update
recorded student model



System presents material and quiz items
based on learner's critical concepts

A Learner Model



● Assimilated Concept ● Critical Unassimilated Concept ● Noncritical Unassimilated Concept

Online Teaching

Question Selection

- Student selects the theme to explore, and is presented some descriptive material.
- When student presses the "Quiz me" button, the system selects the question that tests the most critical concepts for the student.

Question Selection

- Concepts ranked on a combination of:
 - **Relevance:** centralness to the Topic being explored by the student
 - **Criticalness:** a score combining order of difficulty with degree of assimilation (less difficult unassimilated concepts score highest)
 - **Importance:** Frequency of occurrence of errors in relation to the concept (or degree of underuse of the structure)

Learner selects the theme they wish to study
(currently only a Noun Phrase module available)

English Grammar Learning System

Predeterminers



all just of double half almost many some none such

Central Determiners



the another no either that some our this a my both all

Post Determiners



next three last current seven two many first few

Premodifiers



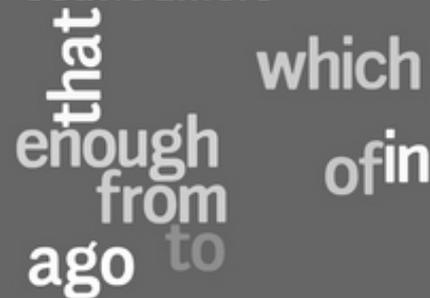
excellent English round smooth paper tiny big small

NP Heads



mine division girl boy handicapped people sheep one costumes car

Postmodifiers



that enough from ago to which of in

Determiners



Article Insertion



Article Choice



Much, Many and Lots



Few, Little



Dual Determiners



Any, All, None, No, Some



Each and Every

Intelligent Grammar Learning System

Indicate which sentence is correct:



Both my twin daughters are coming home for Christmas.



Each my twin daughters are coming home for Christmas.



Either my twin daughters are coming home for Christmas.



Home

Intelligent Grammar Learning System

Indicate which sentence is correct:



Both my twin daughters are coming home for Christmas.



Each my twin daughters are coming home for Christmas.



Either my twin daughters are coming home for Christmas.

Wrong: you should revise the following concepts:

- [Use dual determiners for two items](#): When referring to two entities, use a dual term ('both', 'either', 'neither') if possible.
- [each not dual](#): Each should not be used with two items (use either).
- [each cant be nonpartitive predet](#)



Retry



Continue



Home



Conclusions

The TREACLE Project

- This talk has presented the work in progress concerning the design of an online blended learning i-platform which is aimed at improving the grammatical competence of EFL learners in Spanish universities.
- Our intention is 'targeted' learning: identifying the immediately most critical language concepts needed by the learner and presenting material and exercises aimed at educating the learner in regards to those concepts.

The TREACLE Project

- By directly addressing identified lacks in the student knowledge, students progress more quickly
 - less time on already known concepts,
 - less time trying to learn material they are not ready for
- Additionally, studies show that when engagement and motivation are high, learning is more effective.
- System will be tested on UAM students in the coming academic year.



Fin

Error Annotation

- Existing EFL materials give far more weight to phenomena which our students do not have problems with, and give little weight to problems that the students do recurrently stumble over.
- 20% of our grammatical errors involve either including an article before a noun when it is contextually inappropriate, or not including it when it should be.
- However, existing EFL materials do not in general spend a great deal of time teaching the rules behind this phenomenon.