Towards Targeted Learning: Mining a Learner Corpus to Support Adaptive Online Learning

Mick O'Donnell
Universidad Autonoma de Madrid

PRELUDE:
WHERE WE ARE COMING FROM

- Our project
- Our Learner Corpus
- What we have studied

The TREACLE Project

Project: TREACLE

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 Politécnica de Valencia

Funded by Ministerio de Ciencia e Innovación 2010-2013
(FFI2009-14436/FILO)

Our Corpus

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 as texts written, to measure proficiency

Our Analysis

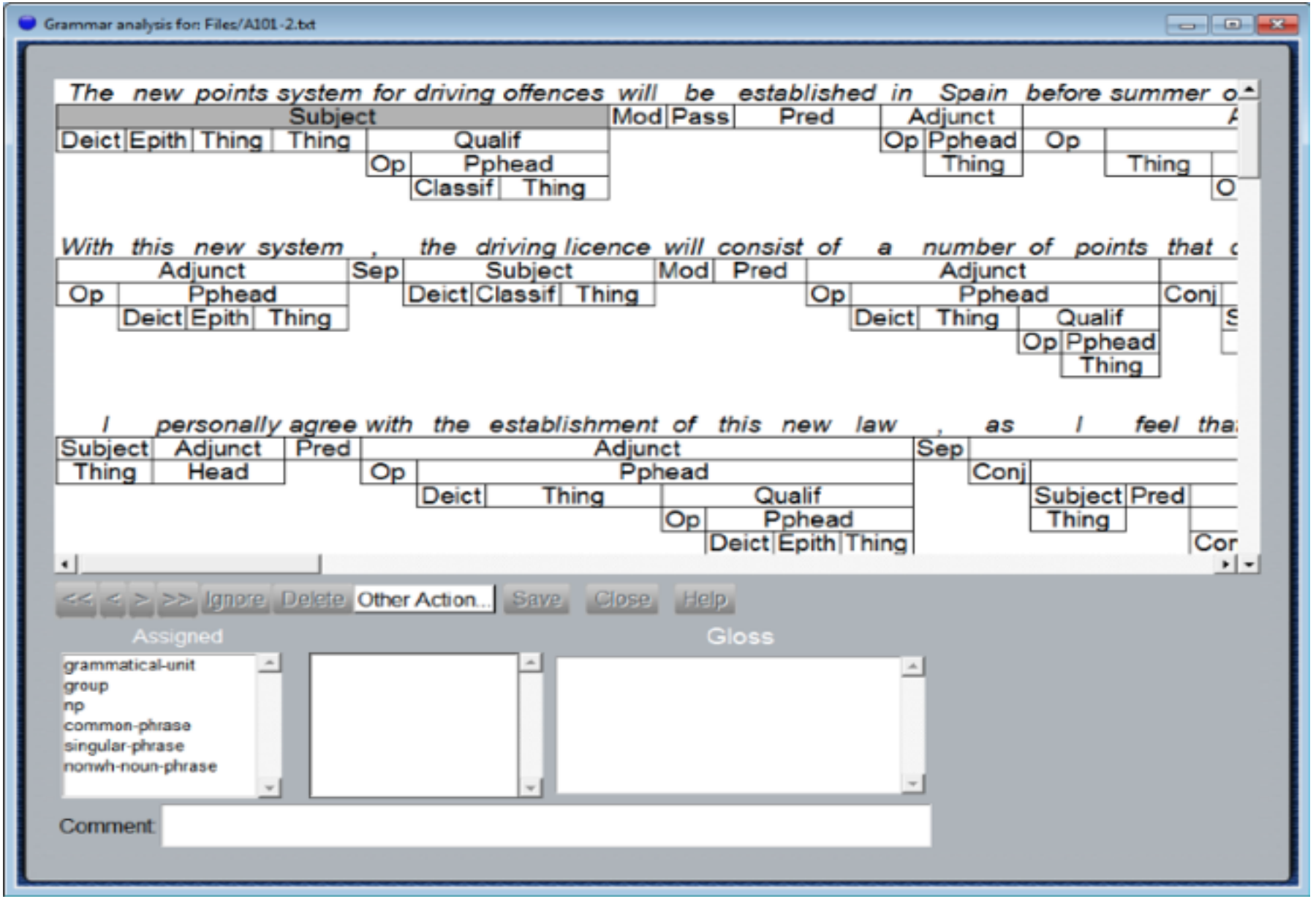
A two-pronged approach was followed:

- 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 analysis of learners' errors
  Coded: 307 essays, 113,000 words, 16,200 errors
  ⟹ *What learners do wrong.*

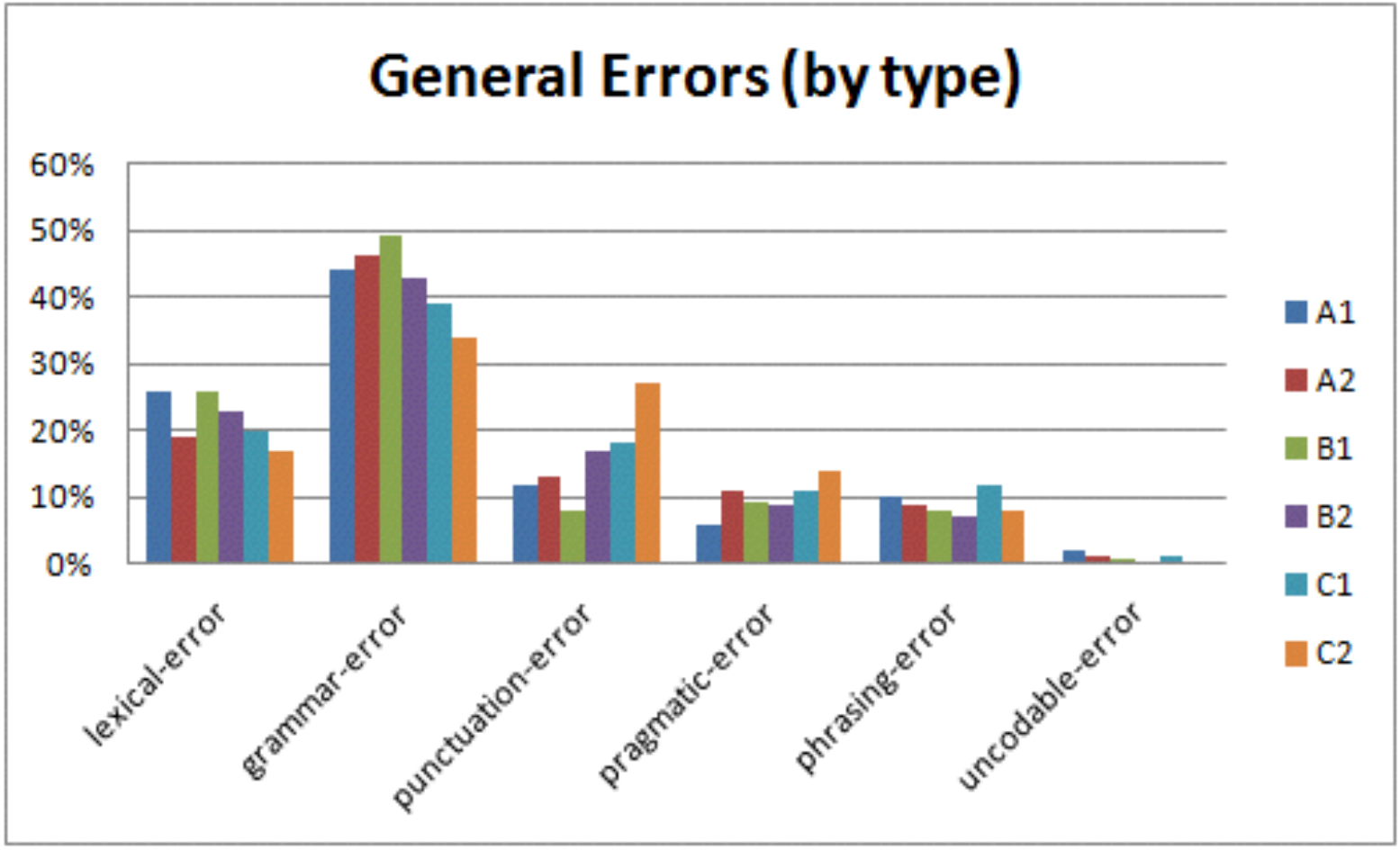
Manual and automatic parsing done via UAM CorpusTool, developed by me, and available from:
http://www.wagsoft.com/CorpusTool/

Automatic Syntactic Annotation

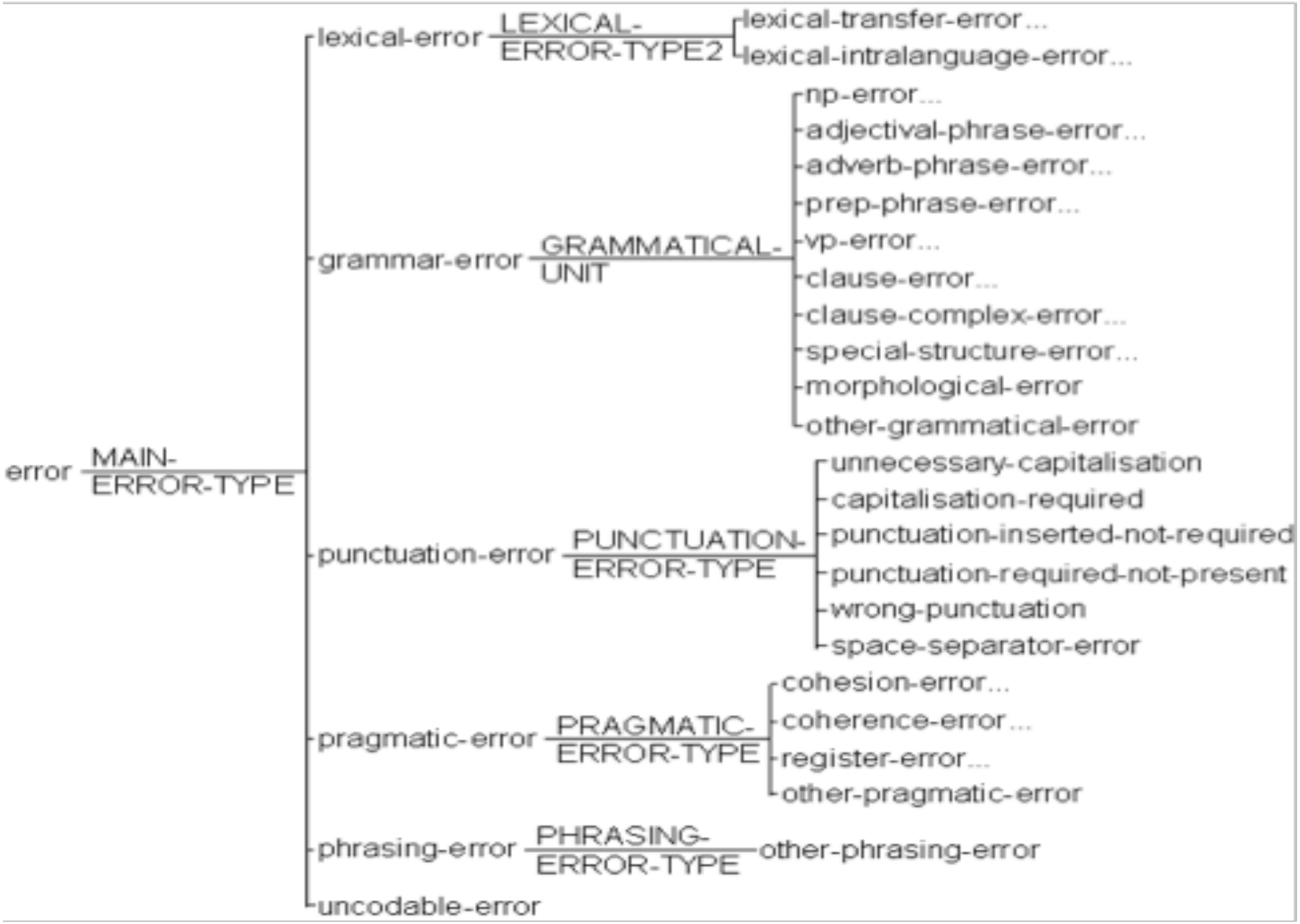


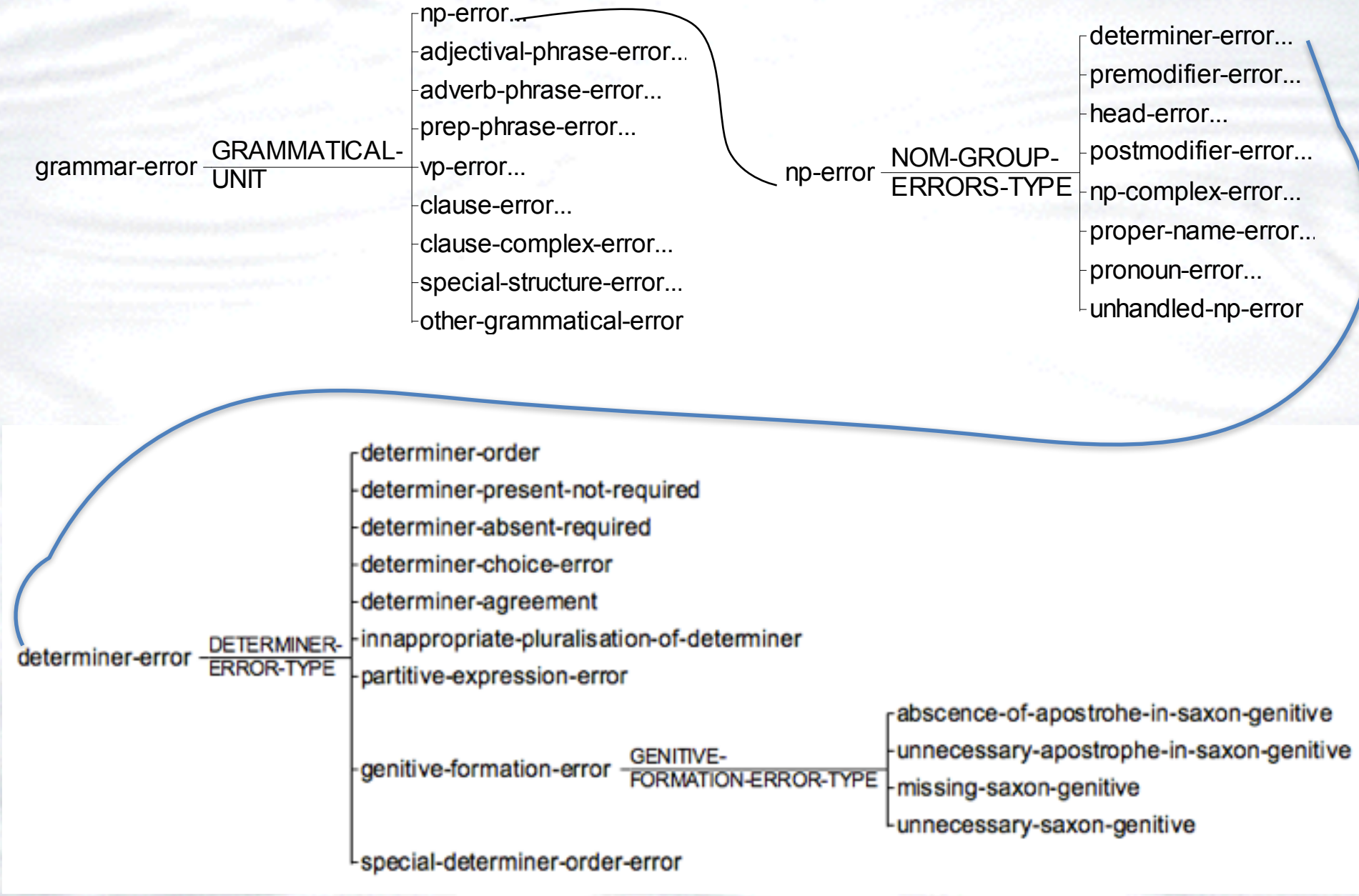
Results of error analysis

Errors per type

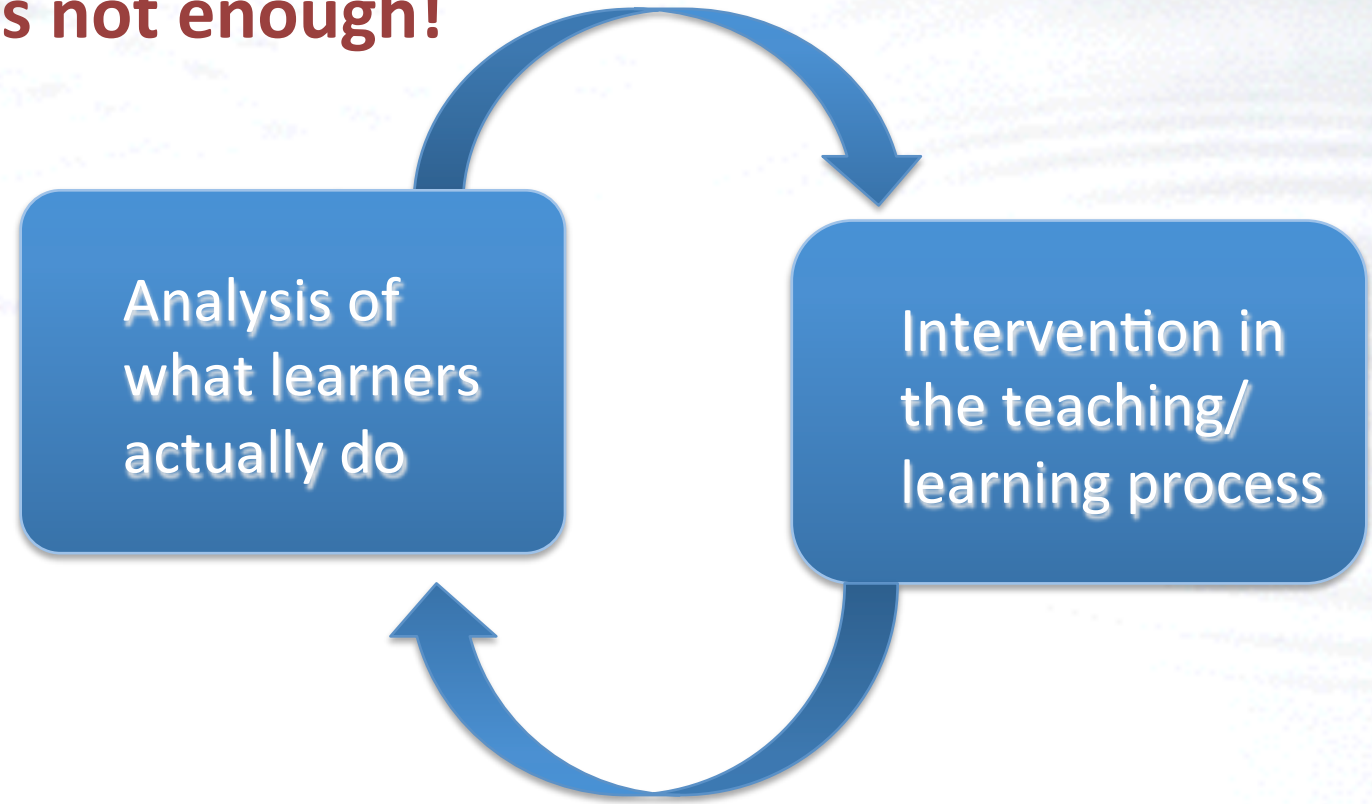


Error Annotation




grammar-error
GRAMMATICAL-UNIT
np-error...
adjectival-phrase-error...
adverb-phrase-error...
prep-phrase-error...
vp-error...
clause-error...
clause-complex-error...
special-structure-error...
other-grammatical-error
np-error
NOM-GROUP-ERRORS-TYPE
determiner-error...
premodifier-error...
head-error...
postmodifier-error...
np-complex-error...
proper-name-error...
pronoun-error...
unhandled-np-error
determiner-error
DETERMINER-ERROR-TYPE
determiner-order
determiner-present-not-required
determiner-absent-required
determiner-choice-error
determiner-agreement
innappropriate-pluralisation-of-determiner
partitive-expression-error
genitive-formation-error
GENITIVE-FORMATION-ERROR-TYPE
abscence-of-apostrohe-in-saxon-genitive
unnecessary-apostrophe-in-saxon-genitive
missing-saxon-genitive
unnecessary-saxon-genitive
special-determiner-order-error

But Analysis is not enough!



- We have developed a large learner corpus, with annotations showing both what learners do at each level, and also what they do wrong.
- But how do we use this information to improve language learning?

PART 1: TOWARDS A TARGETED LEARNING SYSTEM

- Problems with traditional learning
- Targeted learning
- What's needed for a targeted learning system?

Towards Targeted Learning I
The "shotgun" approach to learning

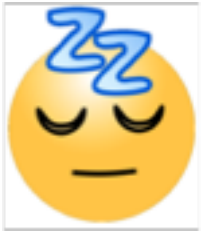
- Many CALL systems take a **shotgun** approach to learning:
  - They have a general idea where the user is,
  - They teach language concepts (grammar, vocab, discourse, etc.) over that area.

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

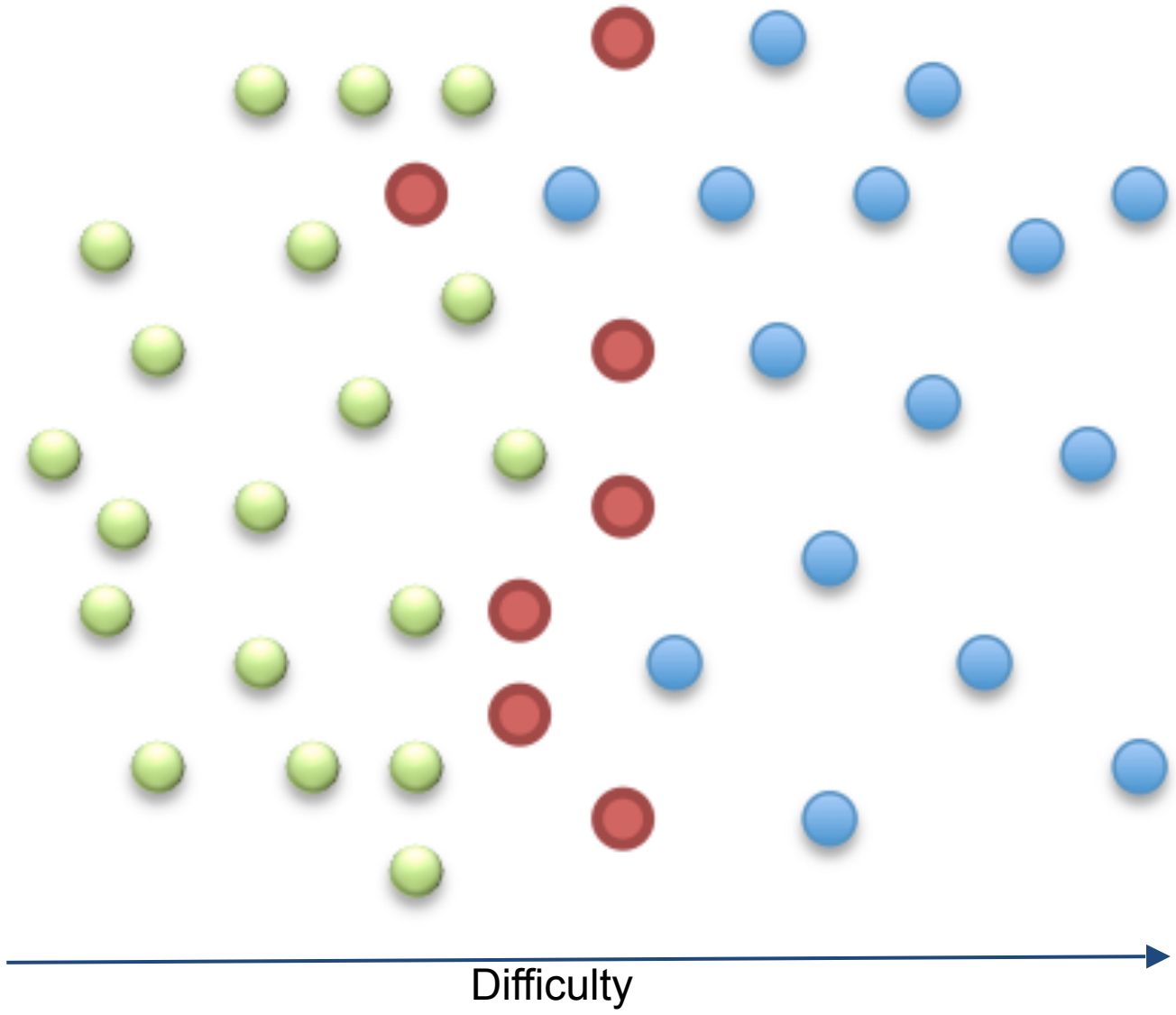
Towards Targeted Learning I

A Targeted Learning Approach

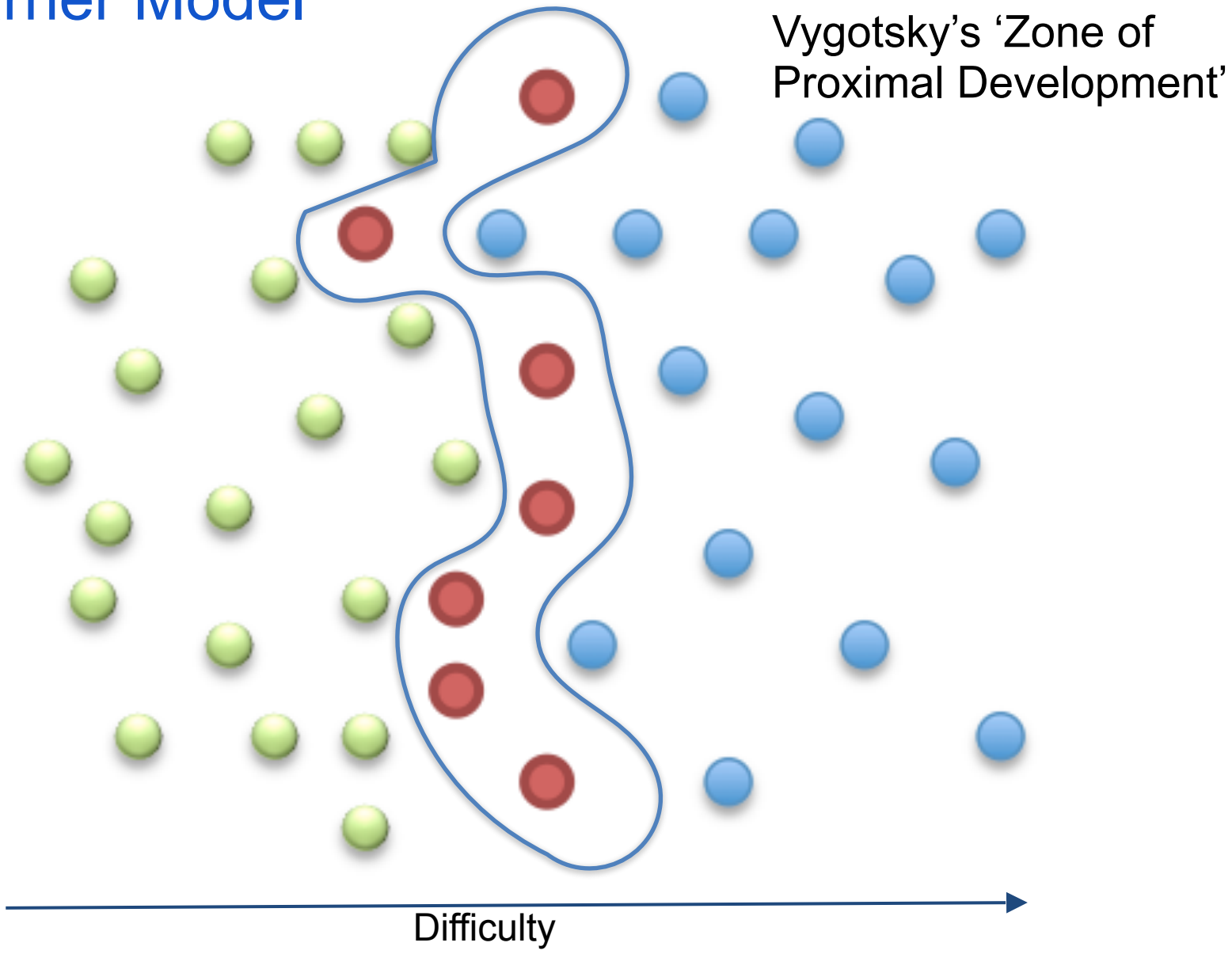
Solution: **targeted learning**: the online system should teach exactly those language concepts that are most critical to the current learner at this point of time.

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

A Learner Model



A Learner Model



Towards Targeted Learning I

A Targeted Learning Approach

Targeting concepts: two aspects to judge importance:

1. Language concepts that are real and observable problems to language learners -> Criticality

2. Language concepts which peers of the learner know but the learner does not -> Timeliness

PART 2:
RESOURCES TO SUPPORT TARGETED LEARNING: TIMELINESS

Timeliness: discovering WHEN to teach concepts

Calculating Timeliness (Approach 1):

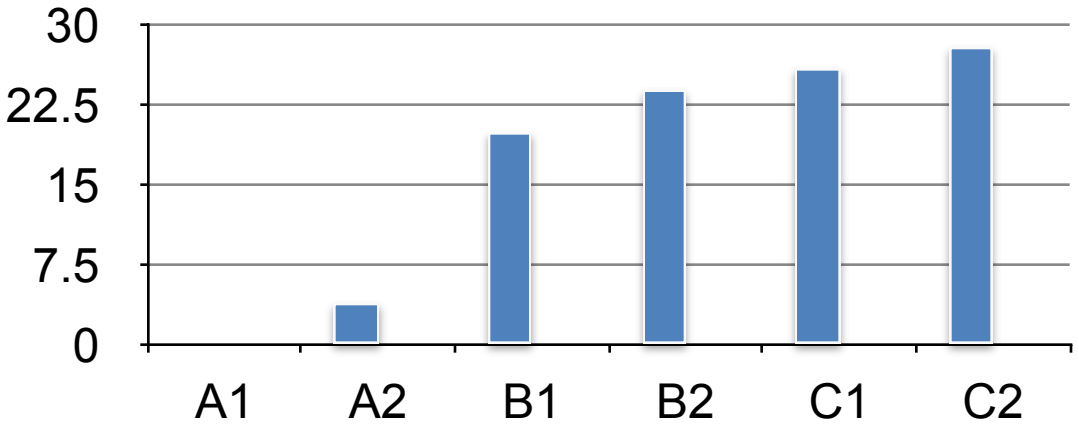
1. Place each grammatical concept at a particular proficiency level.
2. Place each student at a particular proficiency level.
3. Timely concepts are those concepts at the student's level that are not yet acquired.

Timeliness: discovering WHEN to teach concepts

Calculating timeliness (Approach 1):

1. Place each grammatical concept at a particular proficiency level.

- The Cambridge group (Hawkins et al) actually try and do this.
- They claim (using Cambridge Proficiency exams) that there are clear levels where students start to use particular structures:

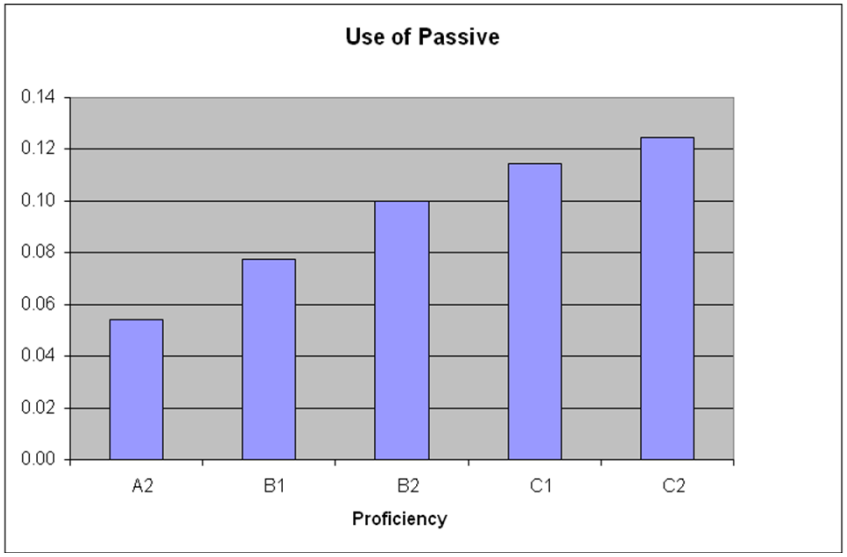


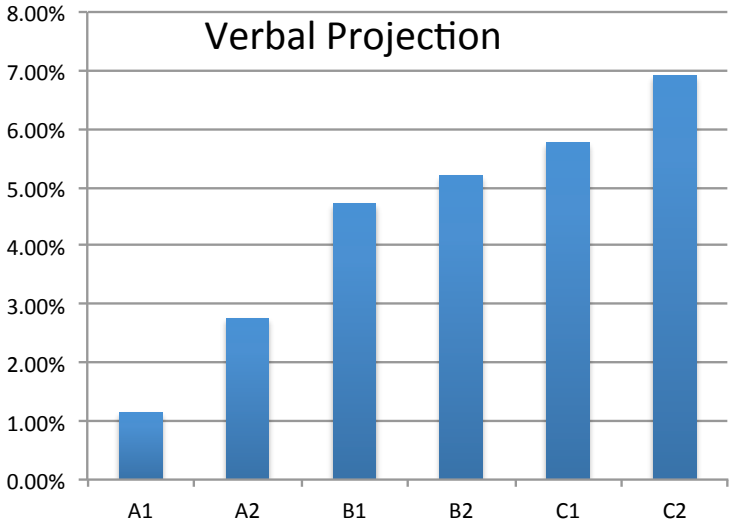
Timeliness: discovering WHEN to teach concepts

Calculating timeliness (Approach 1):

1. Place each grammatical concept at a particular proficiency level.

- But in our learner data, we never see a clear leap from one level to another.
- Rather, it is a continuous improvement over time.
- Where does one decide that the concept belongs?





Timeliness: discovering WHEN to teach concepts

Calculating timeliness (Approach 2):

1. Order grammatical concepts relative to each other in difficulty.
2. Identify degree to which student has mastered each concept
3. Timely concepts are then those concepts lowest in difficulty that the student has not yet acquired.

Timeliness: discovering WHEN to teach concepts

Calculating timeliness (Approach 2):

A lot easier than placing features at proficiency levels

1. Order grammatical concepts relative to each other in difficulty.

2. Identify degree to which student has mastered each concept

3. Timely concepts are then those concepts lowest in difficulty that the student has not yet acquired.

Timeliness: discovering WHEN to teach concepts

Calculating timeliness (Approach 2):

1. Order grammatical concepts re
   other in difficulty.
2. Identify degree to which student has mastered each concept
3. Timely concepts are then those concepts lowest in difficulty that the student has not yet acquired.

Use of quizzes to identify mastery of concepts (or lack of)


Bookmark

Towards Targeted Learning I

A Targeted Learning Approach

1. Timely I:
   a. Placement of grammatical concepts at a particular learning level.
   b. Ability to place student at particular learning level
   c. Timely concepts are those at the students level that are not yet acquired.

   b. Ability to identify degree to which student has mastered concepts
   c. Timely concepts are then those concepts lowest in the difficulty list that the student has not yet mastered.

4. How to order features in difficulty

Timeliness: discovering WHEN to teach concepts

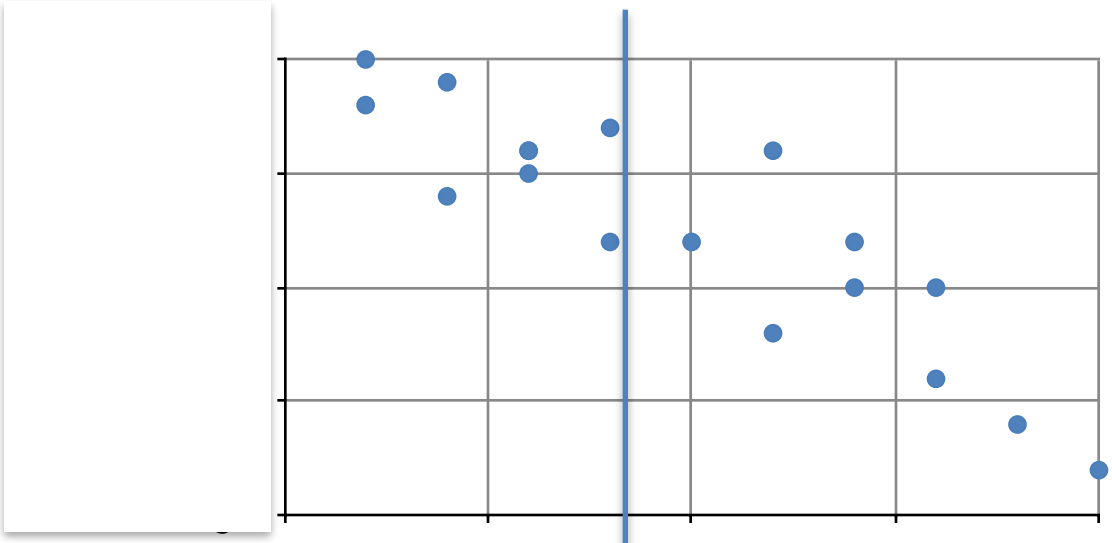
How to order features in difficulty:

Using error data:

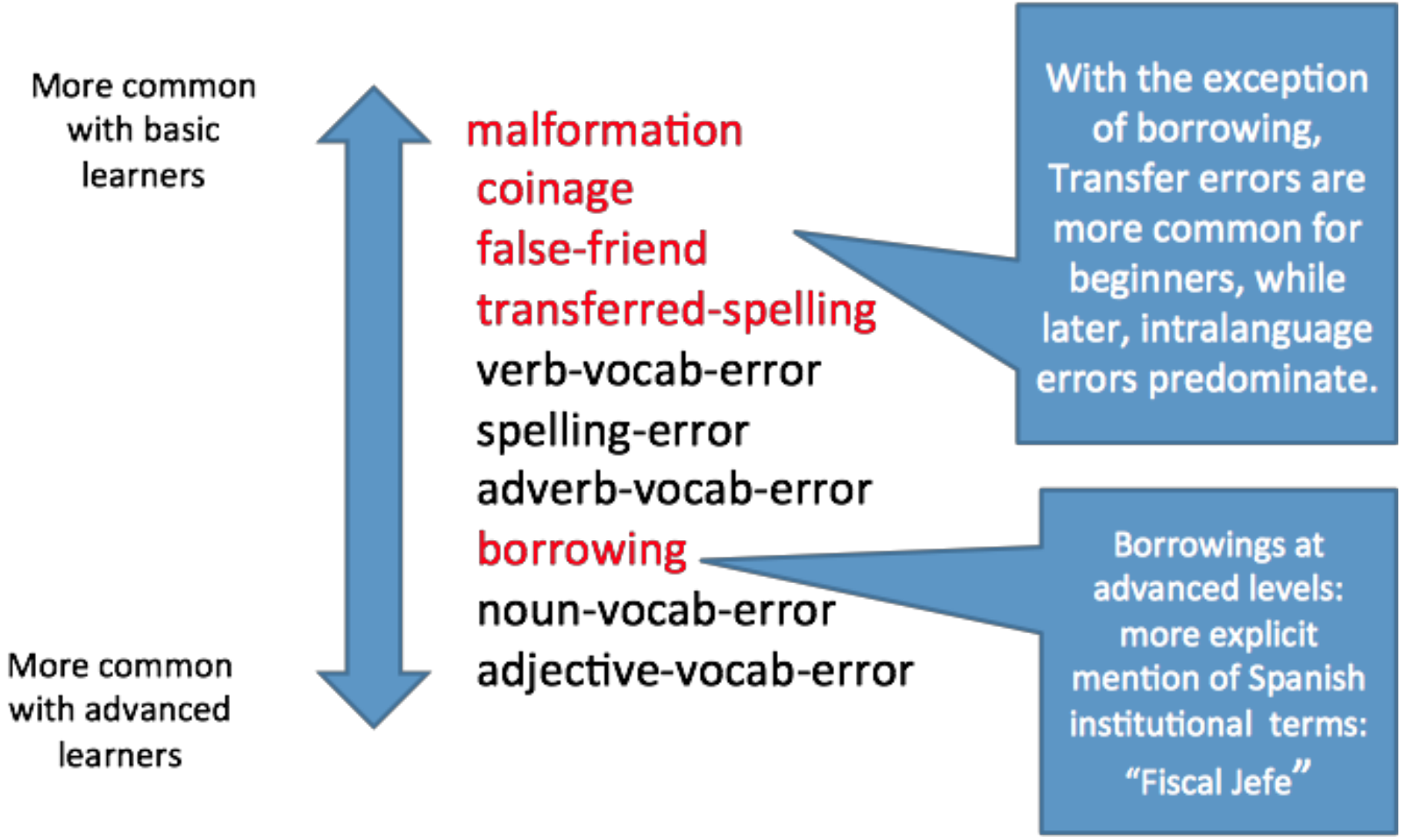
For each error type:

1. Identify each instance of the error
2. Identify the proficiency level of the learner
3. Find average of these proficiency errors

(Errors made more often by low level learners will score lower)



Lexical Errors in terms of apparent difficulty



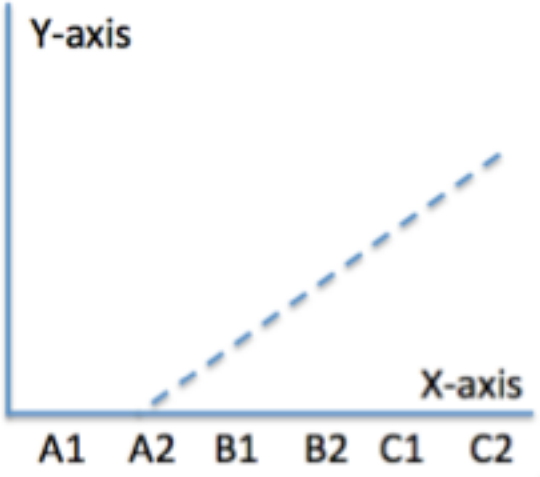
Timeliness: discovering WHEN to teach concepts

How to order features in difficulty:

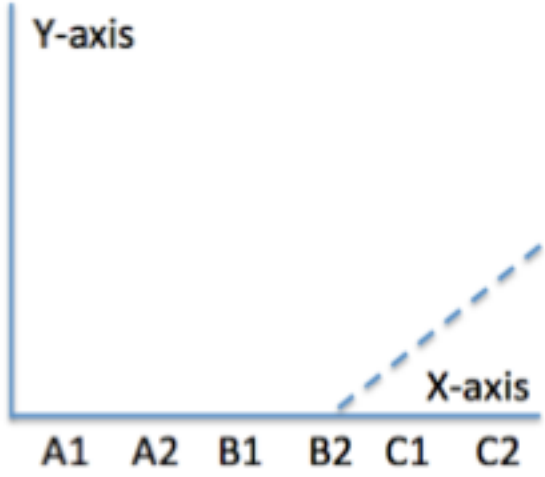
Using syntactic analysis data:

- Various methods, explored in:

Mick O'Donnell (2013) "From Learner Corpora to Curriculum Design: an empirical approach to staging the teaching of grammatical concepts". Proceedings of the V International Conference on Corpus Linguistics (CILC2013). Procedia.



(a) Features with smaller X-intercept are easier to acquire



(b) Features with larger X-intercept are harder to acquire

Timeliness: discovering WHEN to teach concepts

How to order features in difficulty:

Tense-Aspect features ordered in apparent difficulty:

-

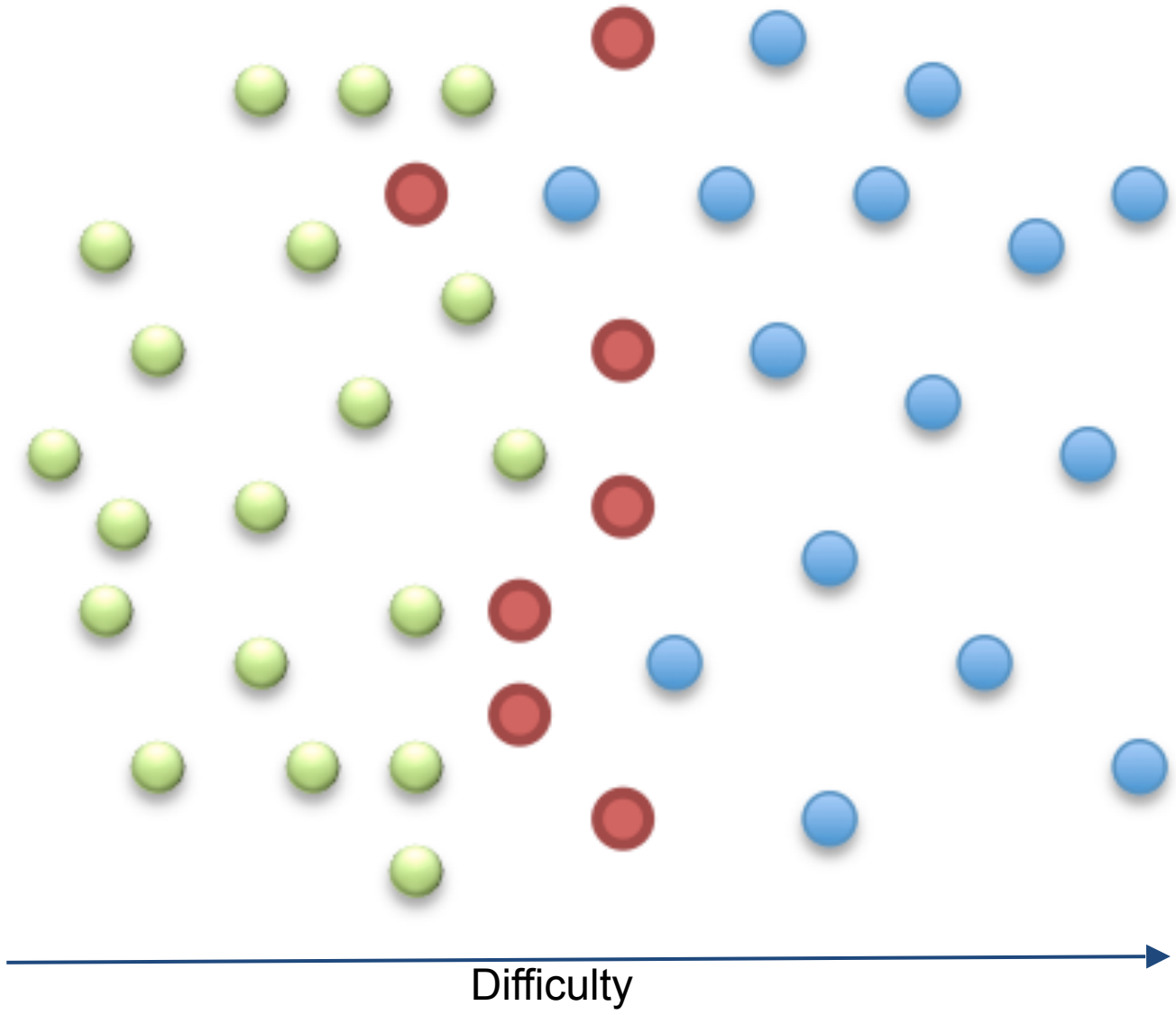
	Y-intercept	relYInterc	Slope
simple-present	0.74068	1.17943	-0.00188
simple-modal	0.12945	0.76097	0.00068
present-progressive	0.03925	1.72916	-0.00028
simple-future	0.03708	1.29066	-0.00014
present-perfect	0.03496	0.57230	0.00044
simple-past	0.01714	0.21332	0.00105
past-progressive	0.00078	0.83713	0.00000
modal-progressive	0.00073	0.66413	0.00001
past-progressive-perfect	0.00045	-5.63573	-0.00001
future-perfect	0.00033	2.13438	0.00000
past-perfect	0.00033	0.10013	0.00005
future-progressive	0.00007	0.14080	0.00001
modal-perfect	-0.00108	-0.51701	0.00005

Timeliness: discovering WHEN to teach concepts

How to order features in difficulty:

- Using these methods, we can order different grammatical features into an ordering of apparent acquisitional difficulty.
- In the online learning system, we use quizzes to discover which grammatical concepts the student has already mastered.
- Those grammatical concepts lowest in the difficulty list which are not acquired are those most ‘timely’ (most within their zone of proximal development)

A Learner Model



Part 2: Resources to support targeted learning: Criticality

Criticality

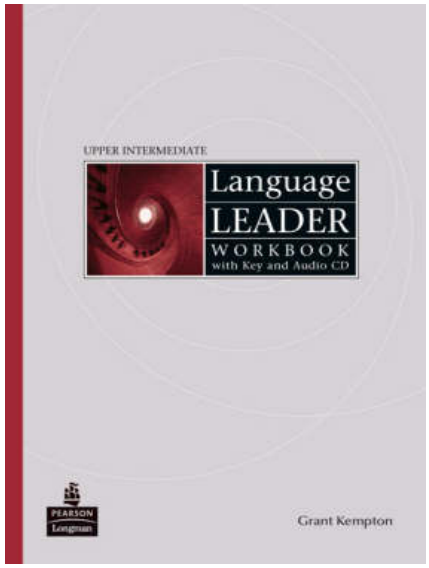
Importance of using criticality in language teaching:

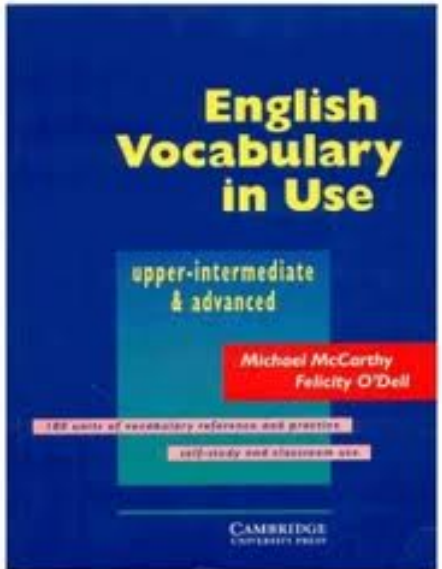
- Learning a foreign language requires mastering thousands of grammatical concepts.
- But many of these concepts are easily acquired, directly transferable from the mother tongue, e.g., English and Spanish share passive structure, progressive aspect, etc.
- So, observing where they go wrong (errors) can help us identify what parts of the language are critical for their learning.

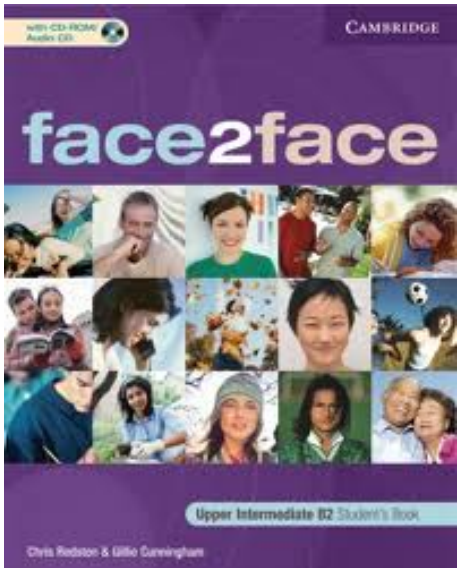
Criticality

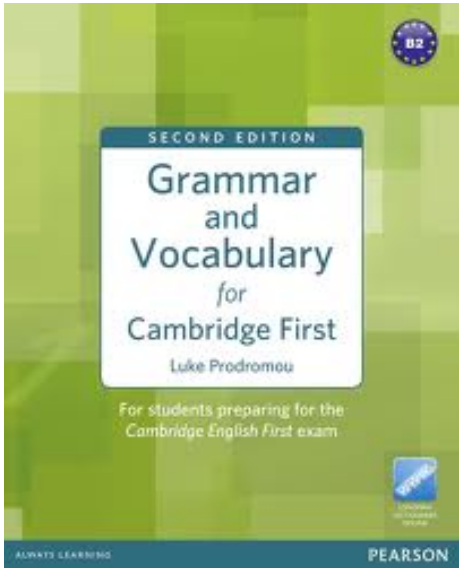
Importance of using criticality in language teaching:

- More teaching effort given to the critical concepts for the particular learner group.
- No need to teach concepts that learners can easily transfer from their mother tongue.









Criticality

Calculating Criticality:

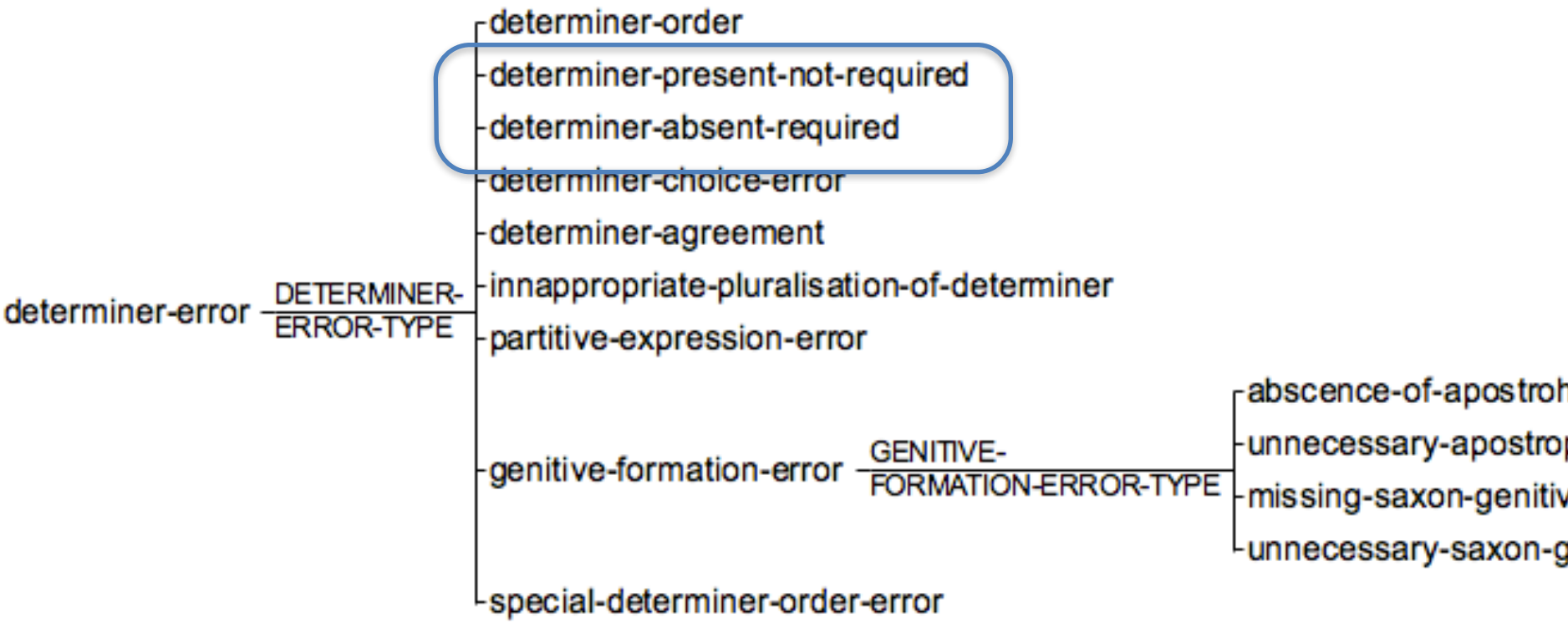
- The most critical concepts are those that learners get wrong most often
- So, relative frequency in our error-annotations identifies criticality.

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

From error-tag to grammatical concept:

- Most of our error codes do not provide enough detail of the grammatical concept involved.
- We are in the process of refining them in the corpus to reflect “concept broken”



From error-tag to grammatical concept:

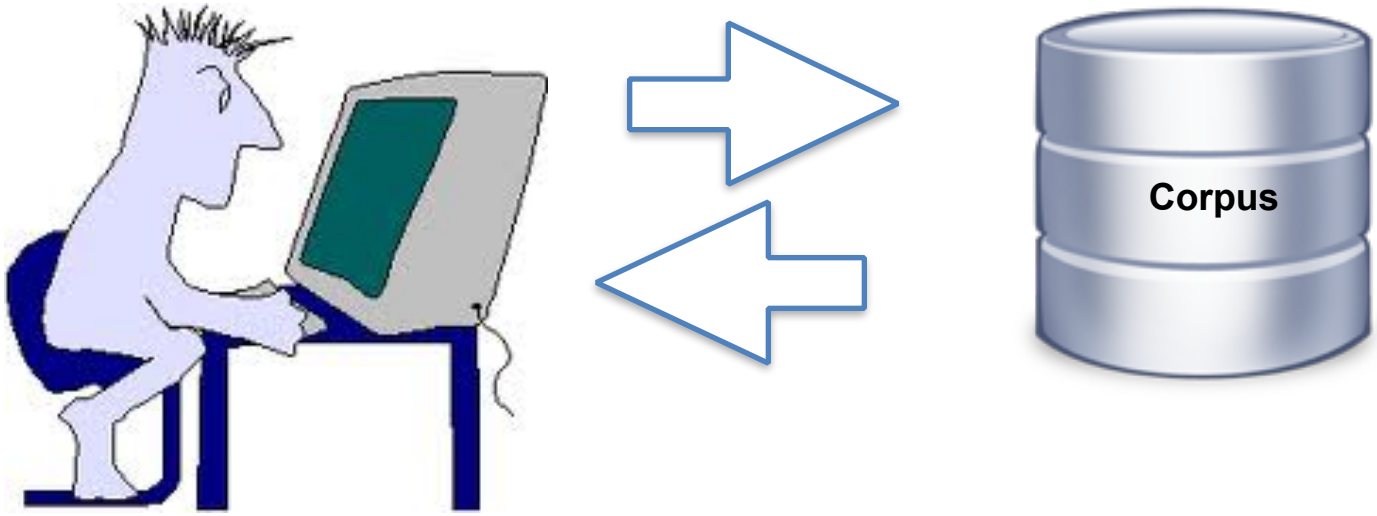
Summary

So, we can derive from our learner corpus the resources we need:

- A ranking of grammatical concepts in terms of acquisitional difficulty (to calculate timeliness)
- A ranking of grammatical concepts in terms of overall frequency of occurrence (criticality)

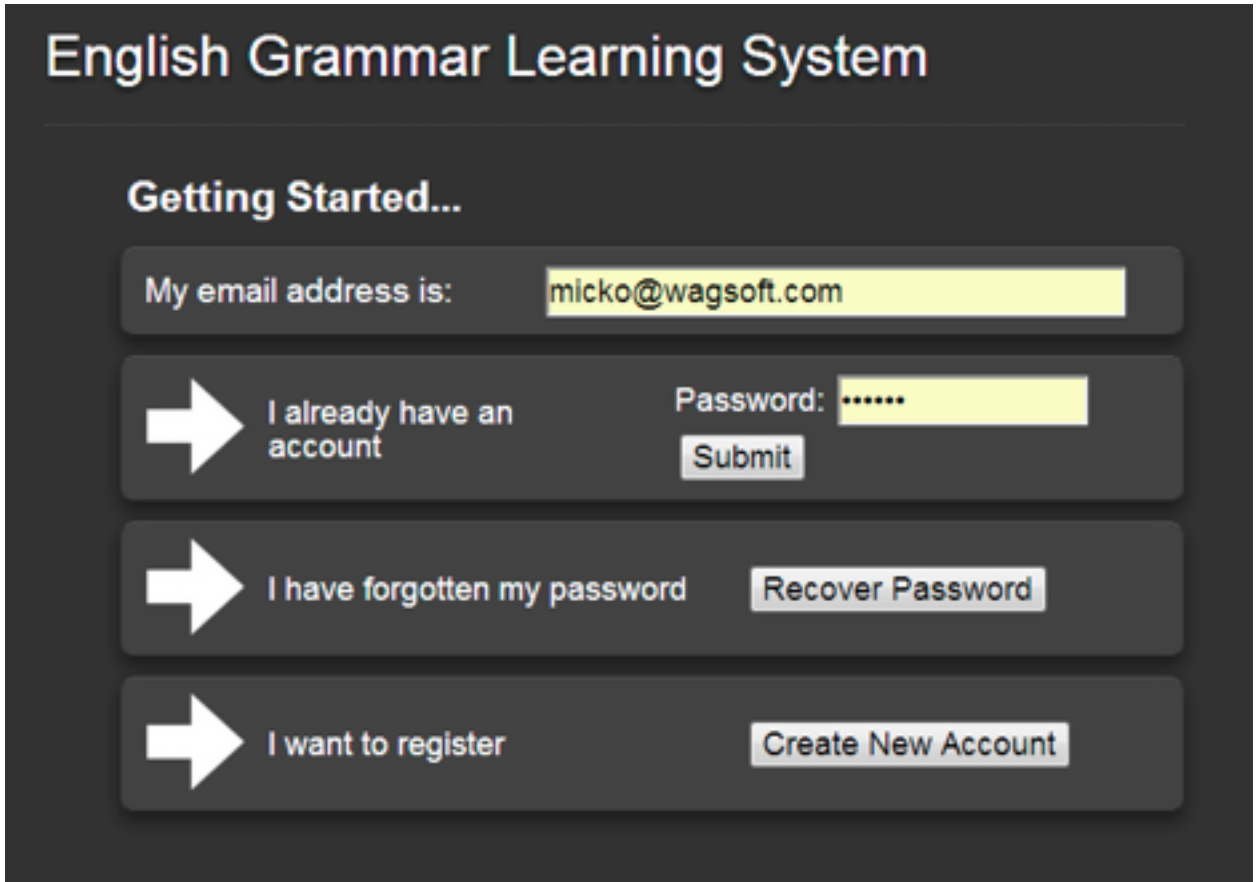
How do we use this information in an online learning system?

PART 3:
PUTTING IT ALL TOGETHER



Towards Targeted Learning II

- A "test of concept" web-system has been produced.
- Plan to deliver to our first year students for 2014-15



The Language Model

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

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.	

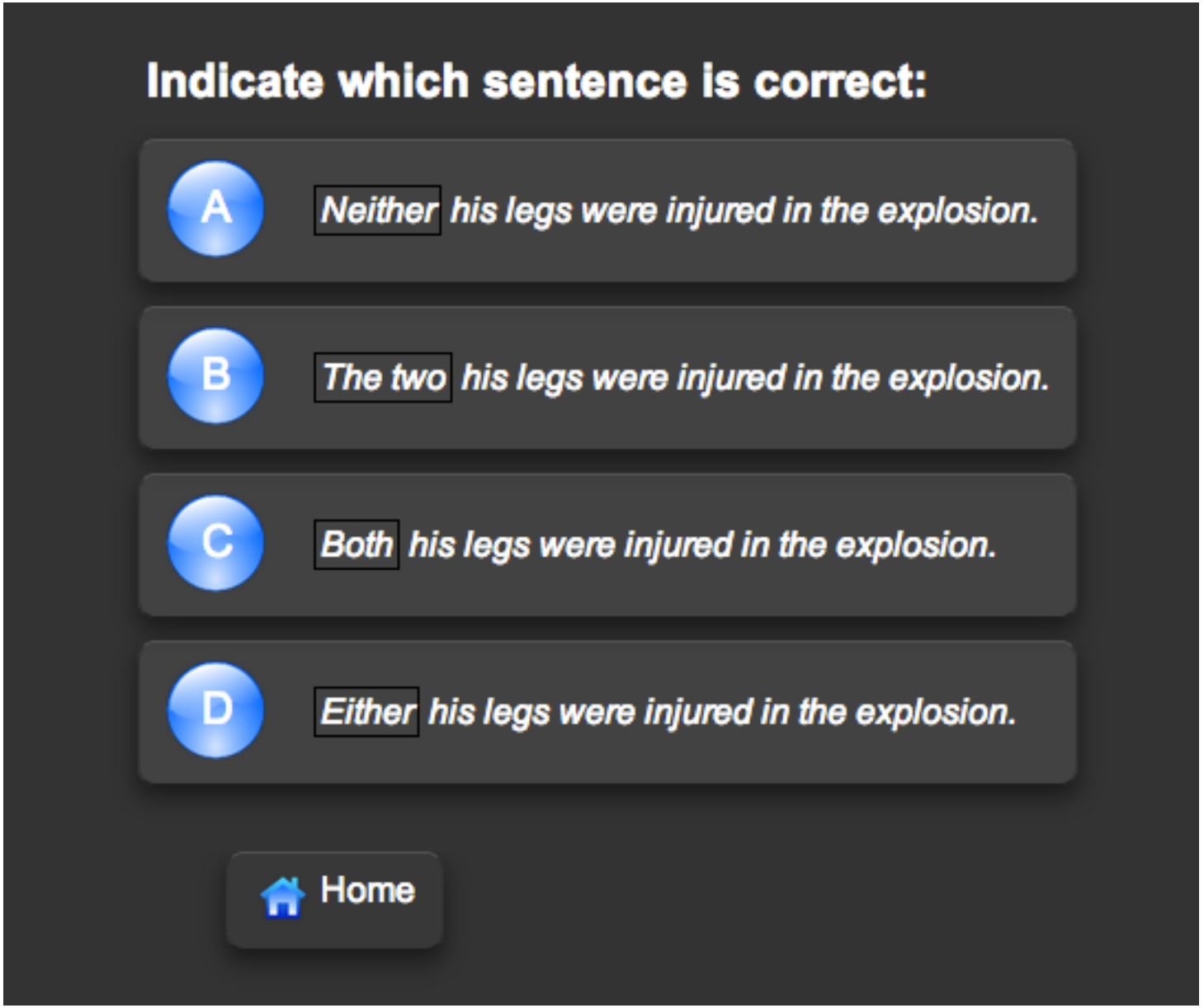
The Question Database

- 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
neither	neither-cant-be-nonpartitive-predet	use-dual-form-where-possible; neither-is-dual
the two	np-cant-be-nonpartitive-predet; use-dual-form-where-possible	
both		both-can-be-nonpartitive-predet; both-is-plural; both-is-dual use-dual-form-where-possible;
either	either-is-singular; either-cant-be-nonpartitive-predet	use-dual-form-where-possible; either-is-dual

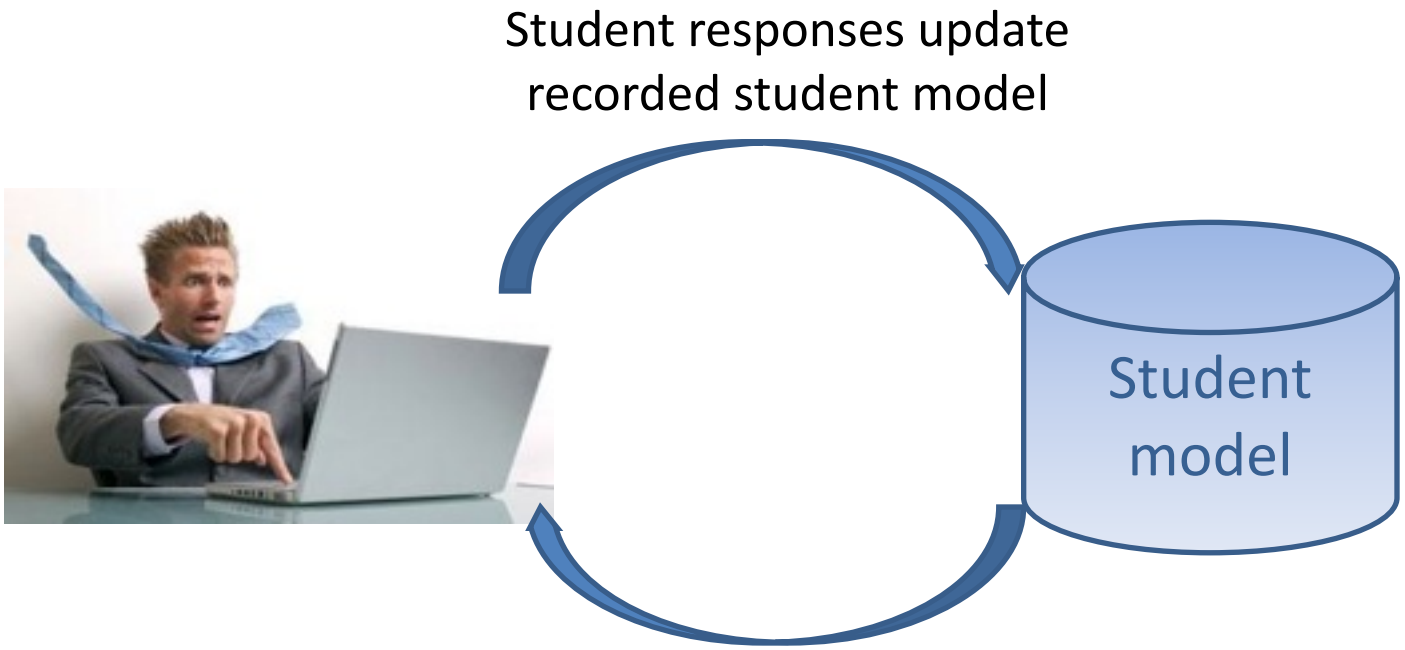
The Question Database



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



Online Teaching

Question Selection

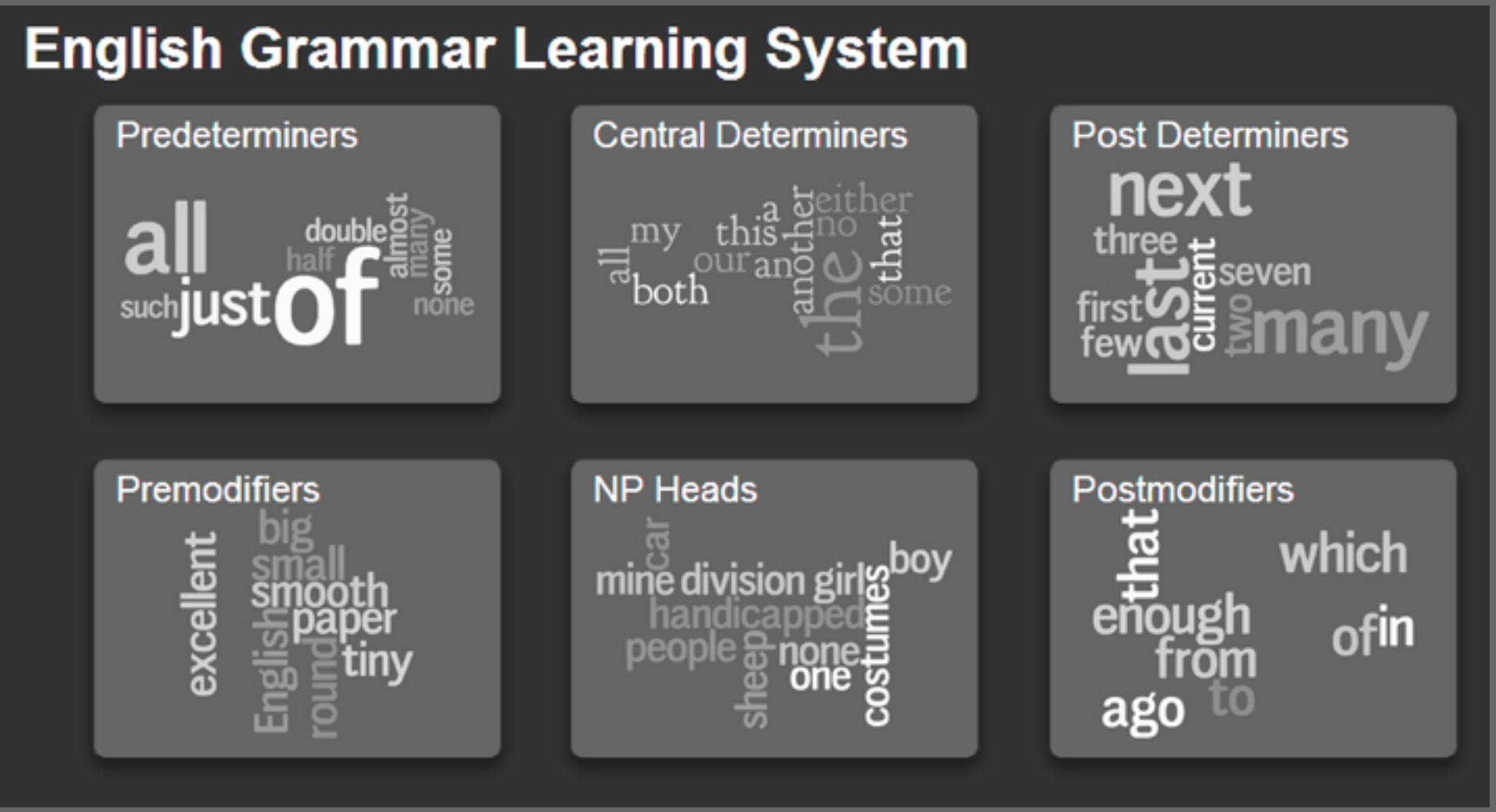
- Student selects the theme to explore, and is presented some descriptive material.

- 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
  - **Timeliness**: a score combining order of difficulty with degree of assimilation (less difficult unassimilated concepts score highest)
  - **Criticality**: Frequency of occurrence of errors in the use of the concept in our learner corpus

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



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:

A *Both my twin daughters are coming home for Christmas.*

B *Each my twin daughters are coming home for Christmas.*

C *Either my twin daughters are coming home for Christmas.*

Home

Intelligent Grammar Learning System

Indicate which sentence is correct:

A *Both my twin daughters are coming home for Christmas.*

Each my twin daughters are coming home for Christmas.

C *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

- To be able to target students with the most relevant learning material to their immediate needs, we derive two kinds of data from our corpus:
  - Grammatical concepts ordered in difficulty
  - Grammatical concepts ordered in terms of frequency of occurrence overall (criticality)
- These are key resources in the adaptive selection of material for the online learner.