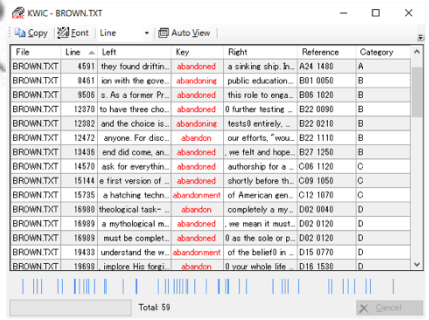
Using a learner corpus to support online intelligent tutoring: the Alegro project



Penny MacDonald
Universitat Politècnica de València

Mick O'Donnell
Universidad Autónoma de Madrid

The ALEGRO Project

Adaptive
Learning of
English
GRammar
Online

A cooperation between:

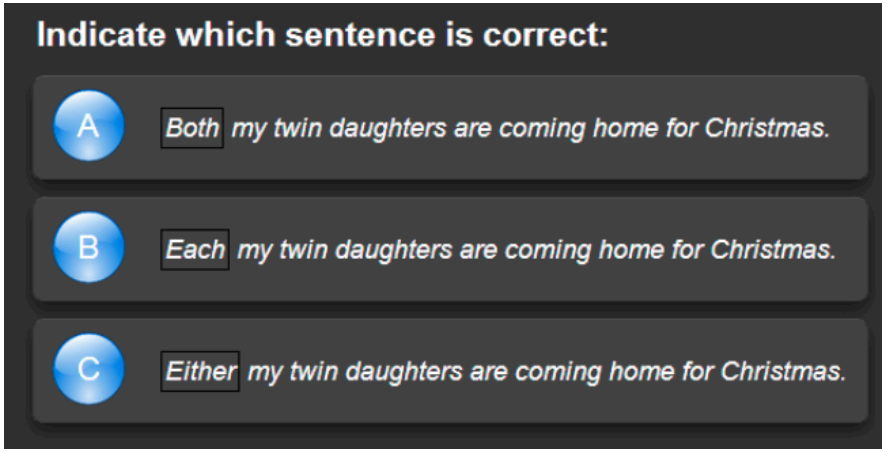
- Universidad Autónoma de Madrid,
- Universitat Politécnica de Valencia
- Universitat de València

Funded by the Ministerio de Economía y Competitividad 2016-2018
(FFI2015-67992-R)

Part 1: An online system for targeted learning of English Grammar

The Alegro System

- The goal of our project is to develop an online learning system to assist our Spanish University students in the acquisition of important grammatical concepts.
    - Learners choose grammatical themes to study (article usage, quantifiers, etc.)
    - Presented with explanations of concepts
    - Can take quizzes on the concepts.
- Key element: system is adaptive: it tracks learner assimilation of concepts via the quizzes and tailors the learner experience on that basis.



The Alegro System

- The system is adaptive in two ways:

1. Critical concepts: Only addresses the 1000 or so grammatical concepts which are most critical for this learner group.

2. Timely concepts: Focus student attention on exactly those grammatical concepts which are within their Zone of Proximal Development (Vygotsky).

The Alegro System

By “grammatical concept” we mean:

- A packet of information a speaker needs to produce the language well.
- (mostly equivalent to “rule”)

E.g.

- `When referring to two items, use a dual determiner.`
- `“both” is a dual determiner.`
- `“both”+NOUN  is plural.`

etc.

Part 2: Deriving critical concepts from a learner corpus

Part 2: Deriving critical concepts from a learner corpus

Critical Language Concepts

- There are 100s of thousands of grammatical concepts (rules, features) that need to be acquired to master a language.
- Many of these can be transferred from the mother tongue.
- Others only infrequently cause problems for the learner.
- It makes sense then to focus on exactly those grammatical concepts which demonstrably cause problems for the learner.

Part 2: Deriving critical concepts from a learner corpus

Critical Language Concepts

- The language concepts of an L2 which are critical differ from one L1 to another.
- We want the system to focus on the 1000 most critical grammar concepts for Spanish University learners of English.
- We can study grammatical errors by this group to identify their critical grammar concepts.

Part 2: Deriving critical concepts from a learner corpus

Deriving the most critical concepts

- Goal: identify the most critical concepts that our language learners need to learn.
- Materials: an annotated error corpus from an earlier project of ours (the TREACLE project), with 16,109 errors identified.
    - Of these 16,200 errors, 7,400 are grammar related.

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

Annotation

A two-pronged approach for tagging the data:

- Manual analysis of learners' errors
  - 307 essays, 113,000 words, 16,109 errors

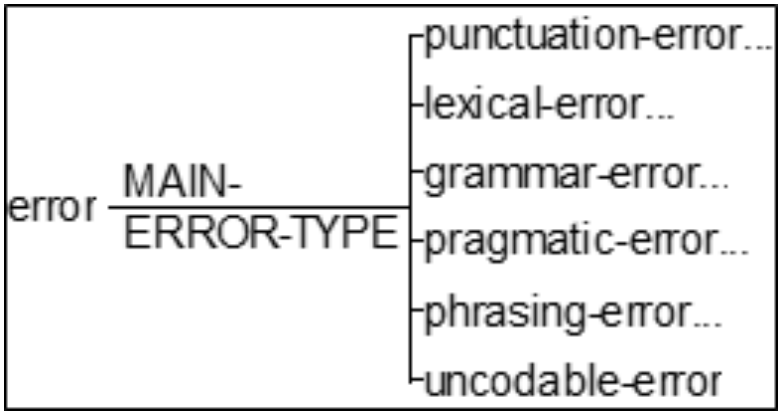
  ⇒ *What learners do wrong.*

- Automatic analysis identifying syntactic structures used by the learners:
  - 1330 essays, 700,000 words, 98,000 clauses

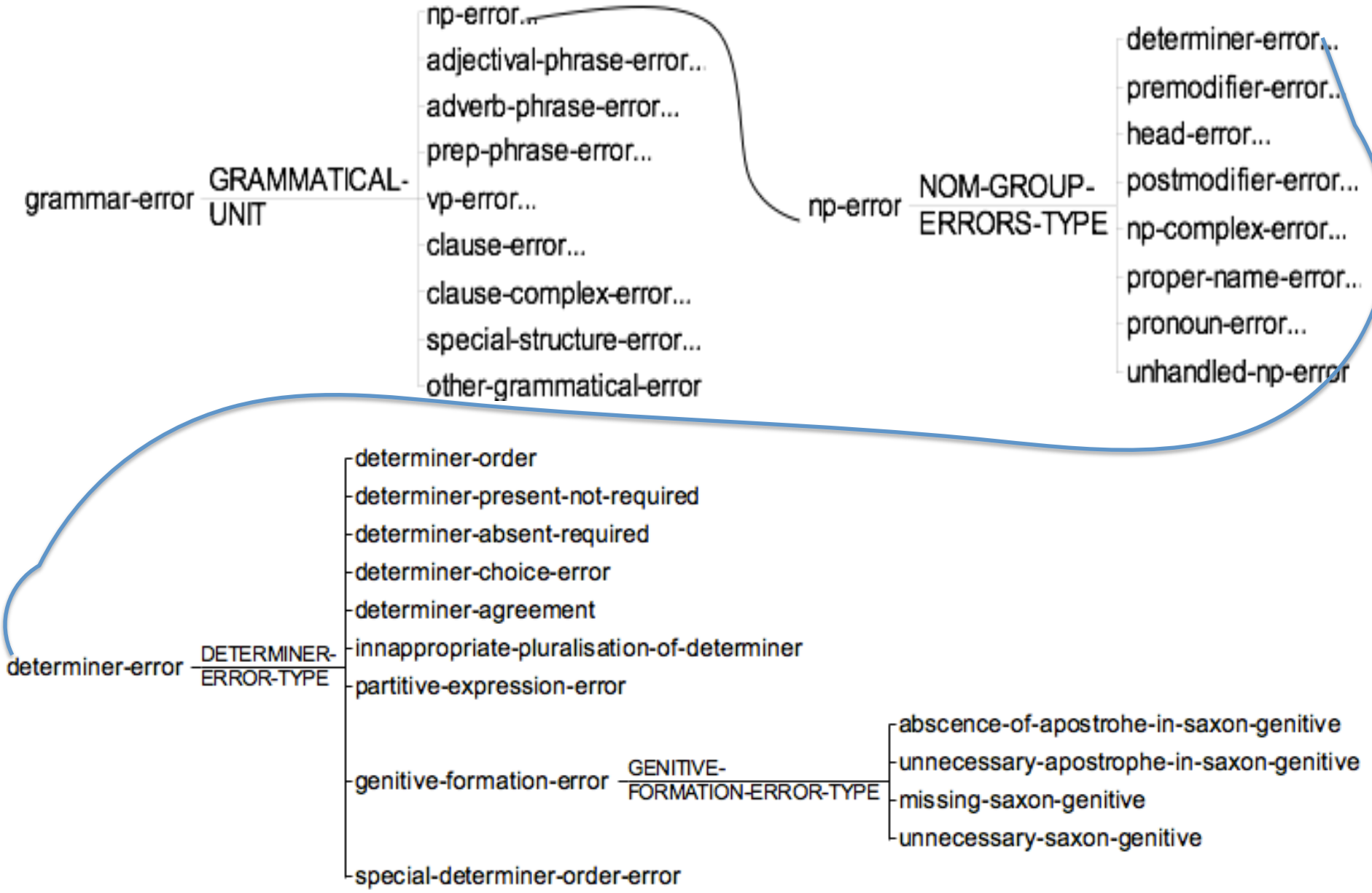
  ⇒ *What learners are doing / not doing*

Manual and automatic annotation done via UAM CorpusTool, available from: http://www.corpustool.com/

Part 2: Deriving critical concepts from a learner corpus



The error coding scheme contains six main categories of errors amounting to 170 error features in total, of which 132 are leaf features (not more delicately specified).


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

Part 2: Deriving critical concepts from a learner corpus

Calculating Criticality:

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

Part 2: Deriving critical concepts from a learner corpus

Deriving the most critical concepts

Methodology:

1. Identify the 20 most frequent grammatical errors.

2. For each error category,

    a. Examine each error instance in turn

    b. Identify the grammatical concept(s) that were not understood to produce that error.

    c. Tag the error with that grammatical concept.

3. Over the corpus, identify the grammatical concepts that most often caused the errors.

Part 2: Deriving critical concepts from a learner corpus

1. Identifying most frequent grammar errors

(top 12 grammar errors organised by topic)

Topic	Error	Count	% (of Gramm. Errors)
Determiner	det-present-not-required	1087	**14.7%**
	det-absent-required	439	**5.9%**
	determiner-choice	250	3.4%
	determiner-agreement	231	3.1%
Head	wrong-number	408	**5.5%**
	pronoun-choice-error	134	1.8%
	wrong-category	122	1.6%
Preposition	preposition-choice	823	**11.1%**
	unnecessary-preposition	205	2.8%
Clause	subject-finite-agreement	536	**7.2%**
	obligatory-subject-absent	227	3.1%
	adjunct-order	179	2.4%

Part 2: Deriving critical concepts from a learner corpus

Deriving the most critical concepts

- Note however that these grammar errors are not by themselves grammatical concepts in our terms.
- Each structural error can result from a range of misunderstood concepts.
- For example: determiner-inserted-not-required:

Error	Broken concept
The terrorism is bad.	Generic noncount don't take article
The cats are mammals.	Generic plurals don't take articles
The seventy percent of...	Percentages don't take article
I study in the university	Places of work/internment don't take article
See you after the coffee	Mealnames don't take article
The most of my friends...	Most as predeterminer doesn't take article

Part 2: Deriving critical concepts from a learner corpus

Results from study by Fiorella Dotti

The use of the article by Spanish learners of English breaks down into several component uses:

- Referring to **specific** entities
  - Normal “the president” / “el presidente”)
  - Percentages: “10 percent” / “el 10 por ciento”
  - Places of work etc.: “go to university” / “ir a la universidad”
  - Meals: “after breakfast” / “después del desayuno”
- Referring to **generic** entities:
  - Count: singular - “the cat” / “el gato”
  - Count: plural - “Cats” / “los gatos”
  - Noncount: - “Love” / “el amor”

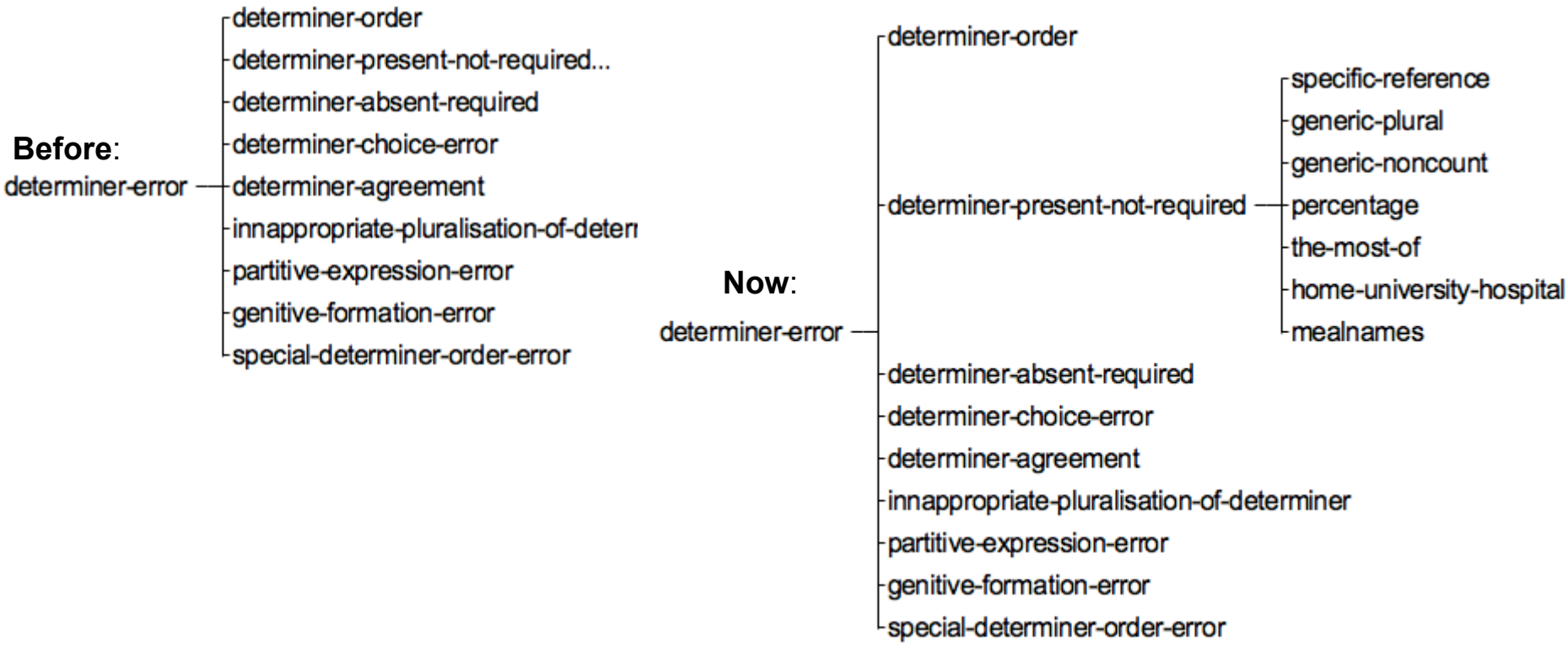
Cases that cause nearly all errors

Part 2: Deriving critical concepts from a learner corpus

Deriving the most critical concepts

So, we need to more finely code our existing error corpus.

- For each error under our most frequent error categories, we need to identify the language concept which was not understood, and thus leads to the error.



Part 2: Deriving critical concepts from a learner corpus

Deriving the most critical grammatical concepts

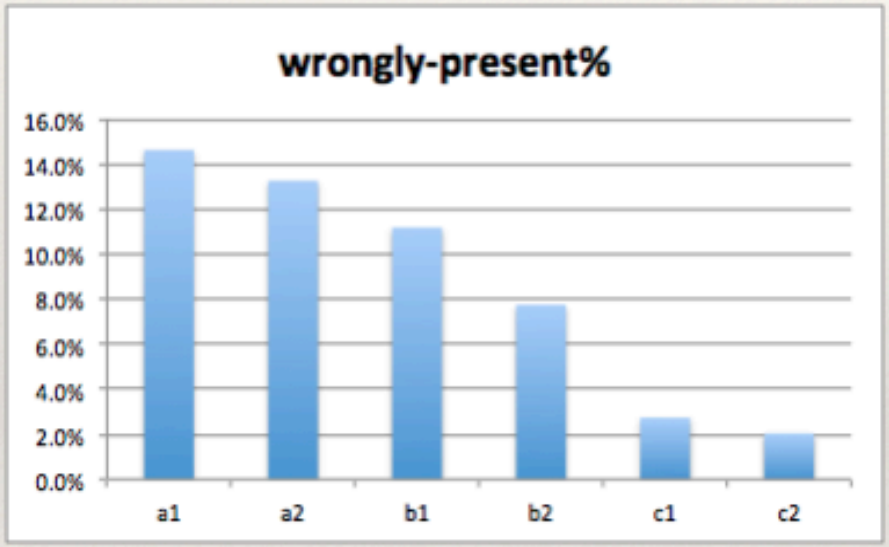
Explaining errors

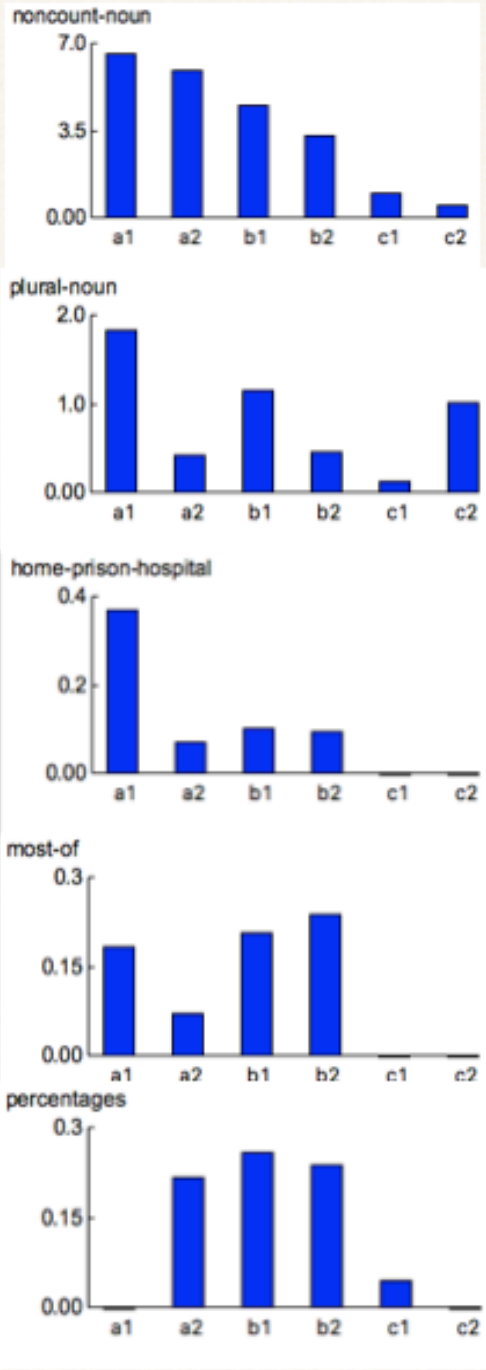
- Often, the learner's misunderstanding is clear from the coder's understanding of both the L1 and the L2.
- We also research language reference books for general explanations of the rules of use of the form in question.

Current Work:

- The work is labor intensive, and our current work involves going through each of the 20 critical errors, identifying underlying cause, and more delicately coding each error.

The more delicate coding of error types allows us to see that what seems to be a smooth progression of development is actually a number of different acquisitional processes working together.





Part 2: Deriving critical concepts from a learner corpus

Some sample explanations (broken concepts)

- **Determiner-head** agreement:
  - *"According to this results"*:
    - `"this" is singular`
      (often not understood because to the Spanish ear, "this" and "these" as essentially the same).
  - "this people"
    - `"people" is plural`
      (the Spanish equivalent, "gente", is singular.

Part 2: Deriving critical concepts from a learner corpus

Some sample explanations (broken concepts)

Results from study by Penny MacDonald and Oksana Polyakova

- **Subject-finite Agreement**:
  - *"People is looking for ..."*:
    - `"people" is plural` ('gente' is singular in Spanish)
  - *"There is reasons ..."*:
    - `Subject in "there" clause follows verb.`

In Spanish,the verb does not change for singular/plural existent:

- *Hay una manzana en la mesa* (There's an apple on the table)
- *Hay dos manzanas en la mesa* (There are two apples on the table)

Part 2.1: Deriving critical concepts from a learner corpus

Subject-finite agreement errors II

- There is a lot of people
- There is always misunderstandings

This can be translated into Spanish as '*hay*' (verb *haber*)

- *Hay una manzana en la mesa* (There's an apple on the table)
- *Hay dos manzanas en la mesa* (There are two apples on the table)

Although syntactically similar, *haber*, in this context does not need to be changed depending on whether it is followed by a singular or plural NP.

Part 2: Deriving critical concepts from a learner corpus

Some sample explanations (broken concepts)

- Quantification:
  - *"too much issues"*:
    - "much" goes with noncount nouns
  - *"I have much time"*
    - Avoid "much" in positive statements.
  - *"I don't have no water"*
    - Avoid double negatives.

Part 2: Deriving critical concepts from a learner corpus

Some sample explanations (broken concepts)

- Prepositions:
  - *"the introduction of tobacco in Europe"*:
    - `use "in" for containment`
    - `use "into" for entering container`
  - *"a picture in the wall"*:
    - `use "on" for noncontained but touching.`

See poster at 17:30 by Patricia Gonzalez on her work on preposition choice errors within this project.

Part 2: Deriving critical concepts from a learner corpus

Summary of Section

- We are in the process of more fine-grained coding of our errors in terms of the grammatical concepts broken in each error.
- From the extended coding, we will be able to identify those grammatical concepts which are most frequently broken.
- The online system will then be given teaching materials to cover these concepts.

Part 3: Keeping learners within their Zone of Proximal Development

Timeliness

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.

Timeliness

Non-adaptive CALL systems

- System teaches all learners a pre-selected set of issues:
  - Some of the material will cover concepts they already know

BOREDOM

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

CONFUSION

Timeliness

Solution: The system keeps track of exactly which grammatical concepts the learner has mastered, those they are still developing, and those they are yet to master.

- The system should then concentrate the learner's online experience within those concepts which they have not yet mastered but are ready to learn

  → Vygotsky's Zone of Proximal Development

- When students are focused within their ZPD, they are maximally engaged (Hamilton & Cherniavsky 2006), and, in such a state of flow, learning is maximised (Csikszentmihalyi 1988)

Timeliness

Building the learner model

- System offers quiz questions to the learner at different points.
- Each answer to the quiz provides evidence that the learner has acquired (or has failed to acquire) different grammatical concepts:

[ ] his legs were injured in the explosion.

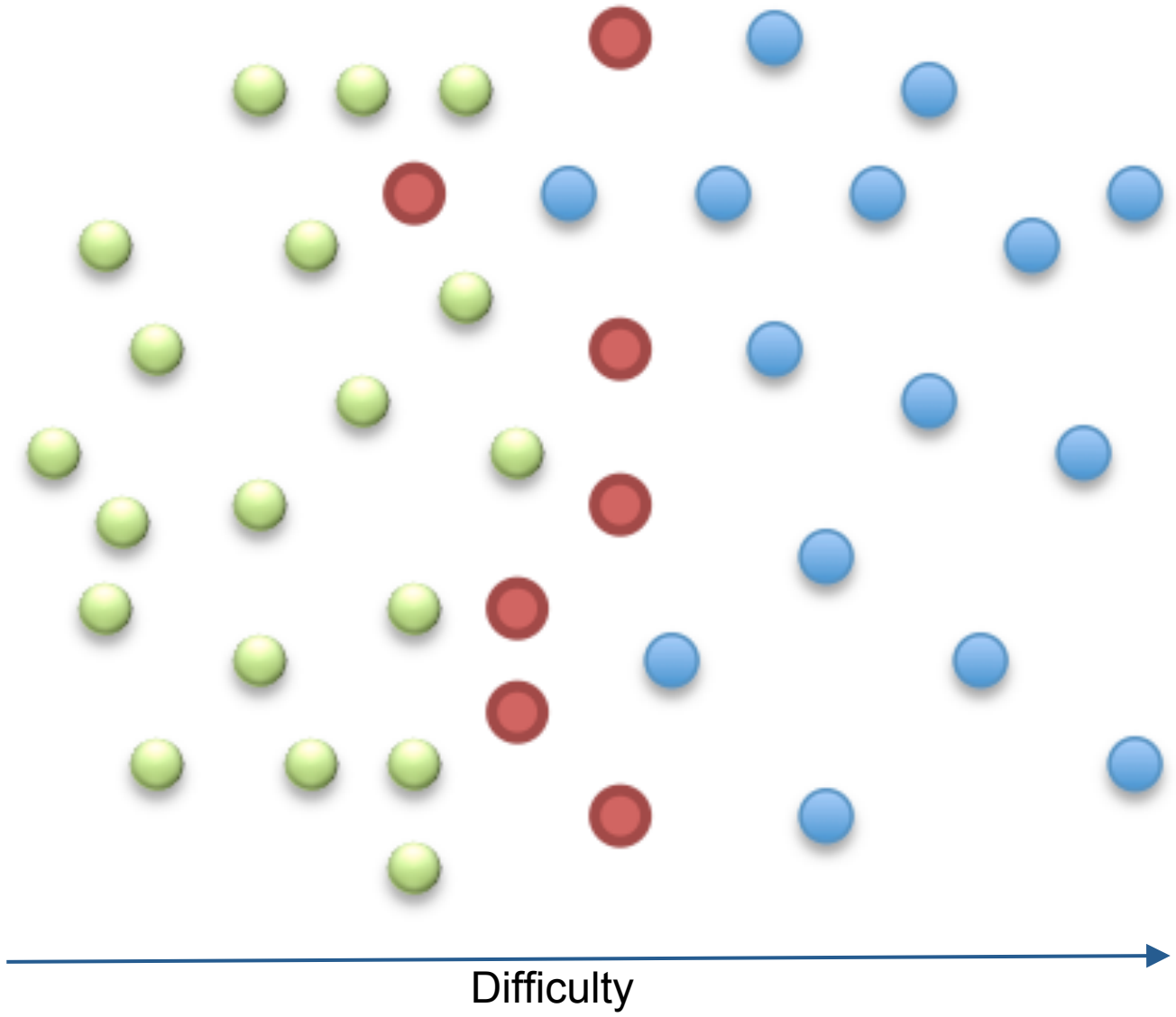
Answer	Concepts Broken	Concepts Complied
neither	neither-cant-be-nonpartitive-predet	use-dual-form-where-possible; neither-is-dual
the two	np-cant-be-nonpartitive-predet; use-dual-form-where-possible	
both		both-can-be-nonpartitive-predet; both-is-plural; both-is-dual use-dual-form-where-possible;
either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

Timeliness

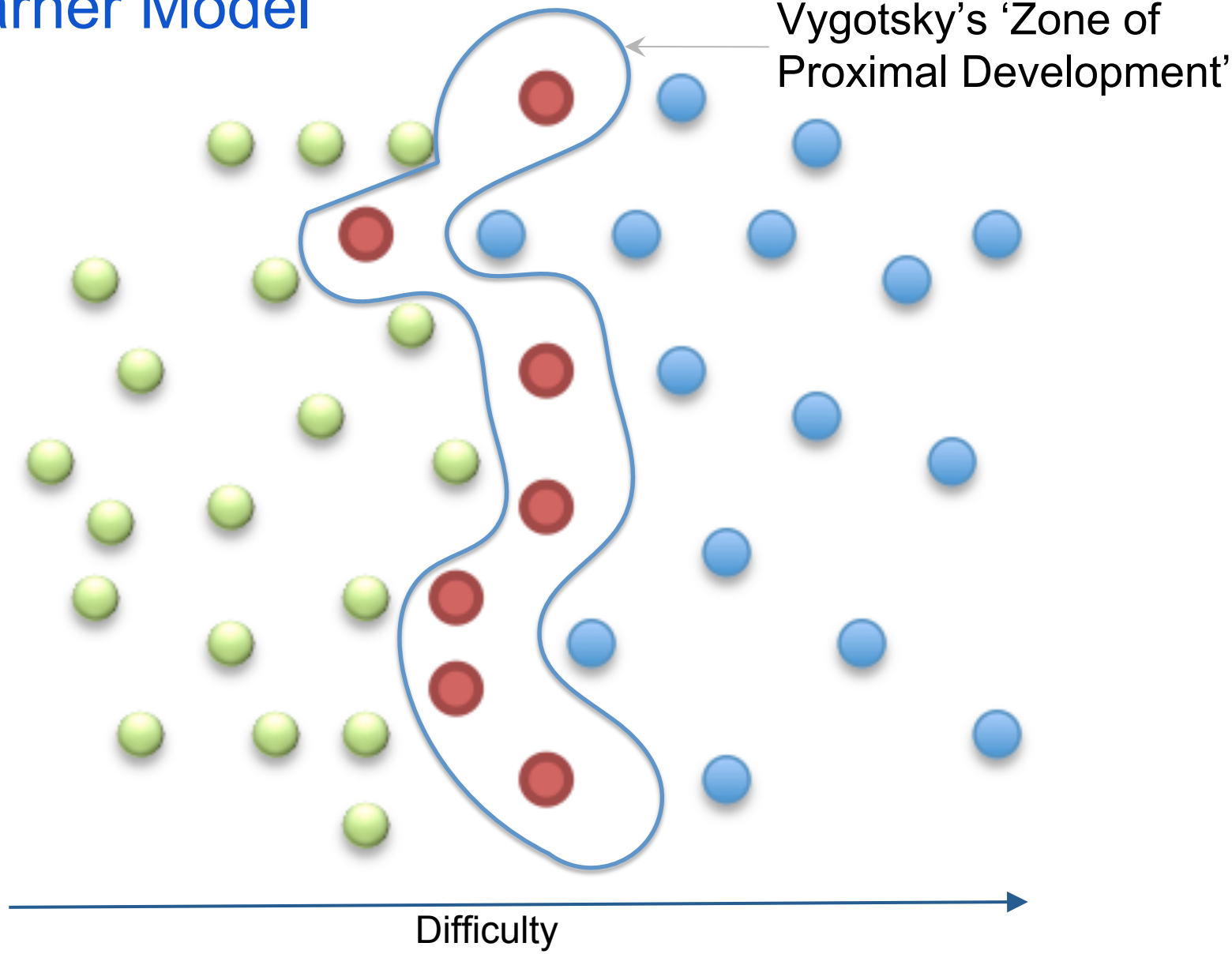
Building the learner model

- Using quiz questions (intermixed with teaching materials), students answer test questions.
- The system builds up a model of the learner's overall grammatical competence.

A Learner Model



A Learner Model



Timeliness: discovering WHEN to teach concepts

Calculating timeliness:

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.

Part 3.2: Deriving the order of developmental difficulty of grammatical concepts from a corpus

Timeliness: discovering WHEN to teach concepts

Calculating Timeliness (Approach 1):

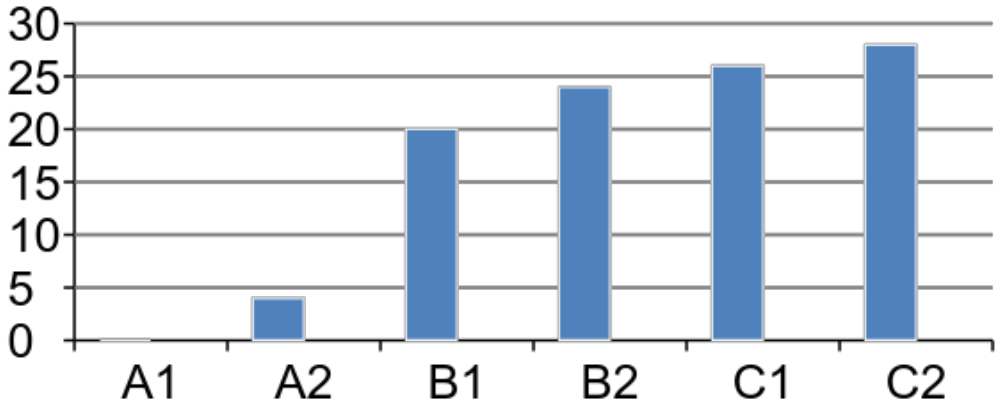
1. Place each grammatical concept at a particular proficiency level.
2. Place each student at a particular proficiency level.
3. Timely concepts are those concepts at the student's level that are not yet acquired.

Timeliness: discovering WHEN to teach concepts

Calculating timeliness (Approach 1):

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

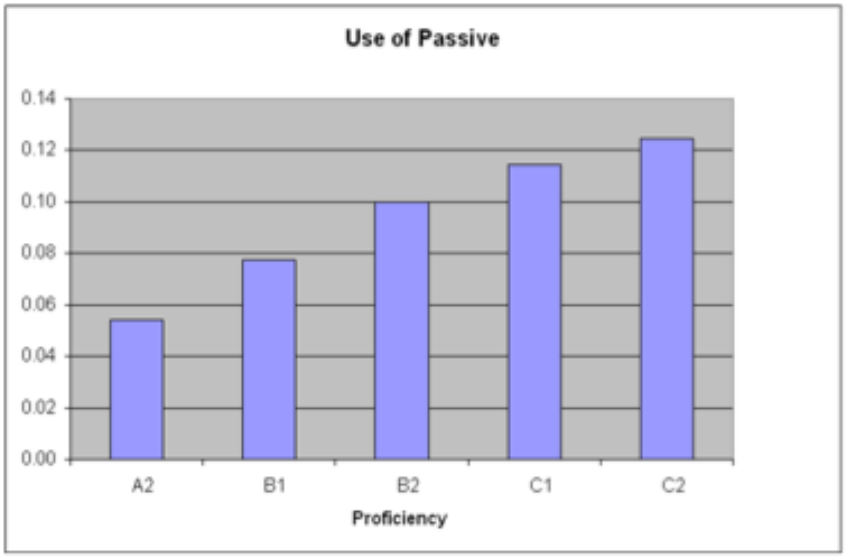
- The Cambridge group (Hawkins et al) take this approach.
- They claim (based on Cambridge Proficiency exams) that there are clear levels where students start to use particular structures:

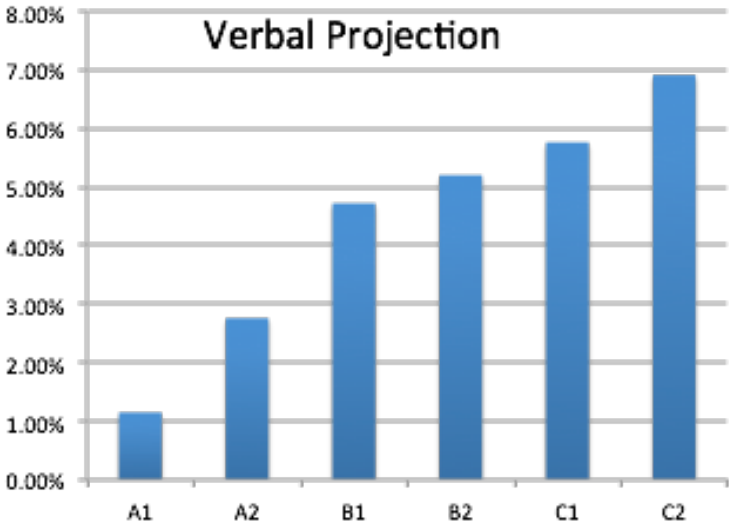


Timeliness: discovering WHEN to teach concepts

Calculating timeliness (Approach 1):

- But in our learner data, we never see a clear leap from one level to another.
- Rather, students exhibit a continuous improvement over time.
- In which CEFR level does does the structure belong?





Timeliness: discovering WHEN to teach concepts

How to order features in difficulty:

Approach 1: Average proficiency level of those who produce errors with the structure

A. For each grammatical concept:

1. Identify essays demonstrating lack of acquisition of the concept.
2. Collect the proficiency levels of these essays
3. Find average of these proficiency scores

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

B. Order the errors from those with lowest average proficiency to those with highest.

Timeliness: discovering WHEN to teach concepts

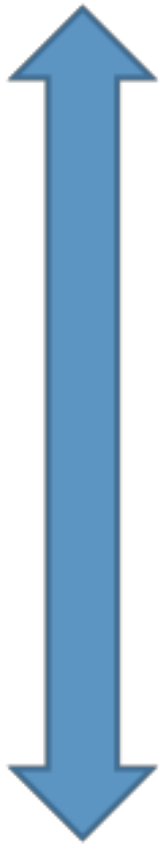
Errors ordered by average proficiency level

other-interrogative-formation-error	28.0
interrogative-formation-error	28.0
incorrect-tense-for-temporal-clause	28.0
ellipsis-error	29.2
pluralised-adjective-head	30.6
postmodifier-order-problem	31.6
adjective-after-head	32.1
intensified-comparative-superlative-adjective	33.0
incorrect-form-for-comparative	33.8

⋮

missing-saxon-genitive	44.2
genitive-formation-error	44.4
wh-nominal-clause-error	44.9
abscence-of-apostrohe-in-saxon-genitive	45.3
incorrect-adjp-complex-connector	45.5
obligatory-object-absent	46.0
wnc-subj-fin-inversion-error	47.2
unnecessary-saxon-genitive	50.5
unnecessary-adverb	54.0

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

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"

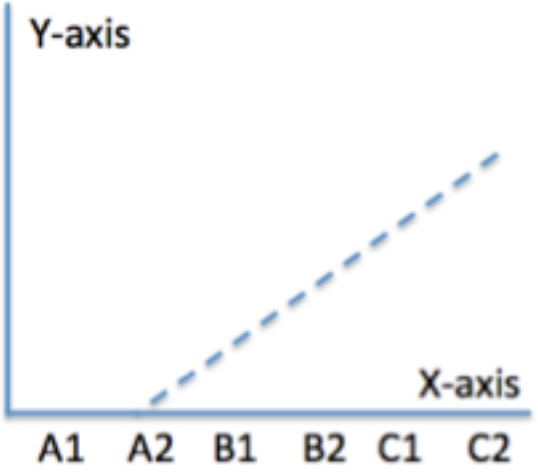
Timeliness: discovering WHEN to teach concepts

How to order features in difficulty:

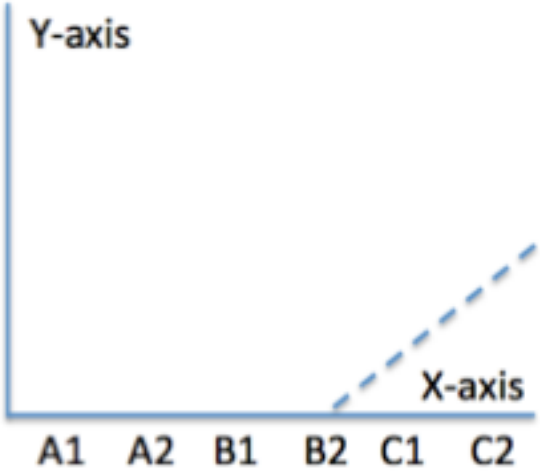
Using syntactic analysis data:

- 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

Problem with using learner corpora for this task

- The lack evidence for A is not evidence for NOT A.
- I some essays, students will not produce a certain error, simply because they don't produce that structure in the essay.
- While "Trust the text" works on large samples, when dealing with small samples, data is too sparse.

Timeliness: discovering WHEN to teach concepts

We thus turned to simple elicitation techniques to measure acquisition of selected concepts:

Each of the following sentences is an attempt to use a question tag. For each of them, indicate whether it is correct or not.

1. You haven't got a car, have you?
   ○ correct ○ wrong
2. He isn't from Germany, he is?
   ○ correct ○ wrong
3. He never came again, did he?
   ○ correct ○ wrong
4. She can't speak Arabic, can she?
   ○ correct ○ wrong
5. I have 5 pets, have I?
   ○ correct ○ wrong
6. Nothing is ready, is it?
   ○ correct ○ wrong
7. He had met him before, didn't he?
   ○ correct ○ wrong
8. You hardly ever came late, didn't you?
   ○ correct ○ wrong
9. They should buy clothes here, shouldn't they?

Timeliness: discovering WHEN to teach concepts

1. Order pairs of concepts: derive tables comparing how often a student demonstrates (non)acquisition of two distinct concepts covered in the test:

		Concept B	
		Not Acquired	Acquired
Concept A	Not Acquired	45	5
	Acquired	15	35

Timeliness: discovering WHEN to teach concepts

2. Ignore cases where no order is indicated

		Concept B	
		Not Acquired	Acquired
Concept A	Not Acquired	45	5
	Acquired	15	35

Timeliness: discovering WHEN to teach concepts

3. Derive order of the concepts

		Concept B	
		Not Acquired	Acquired
Concept A	Not Acquired	45	5
	Acquired	15	35

A < B (A acquired before B)

Timeliness: discovering WHEN to teach concepts

Combining pairwise orderings

$$\frac{A < B \qquad B < C}{A < B < C}$$

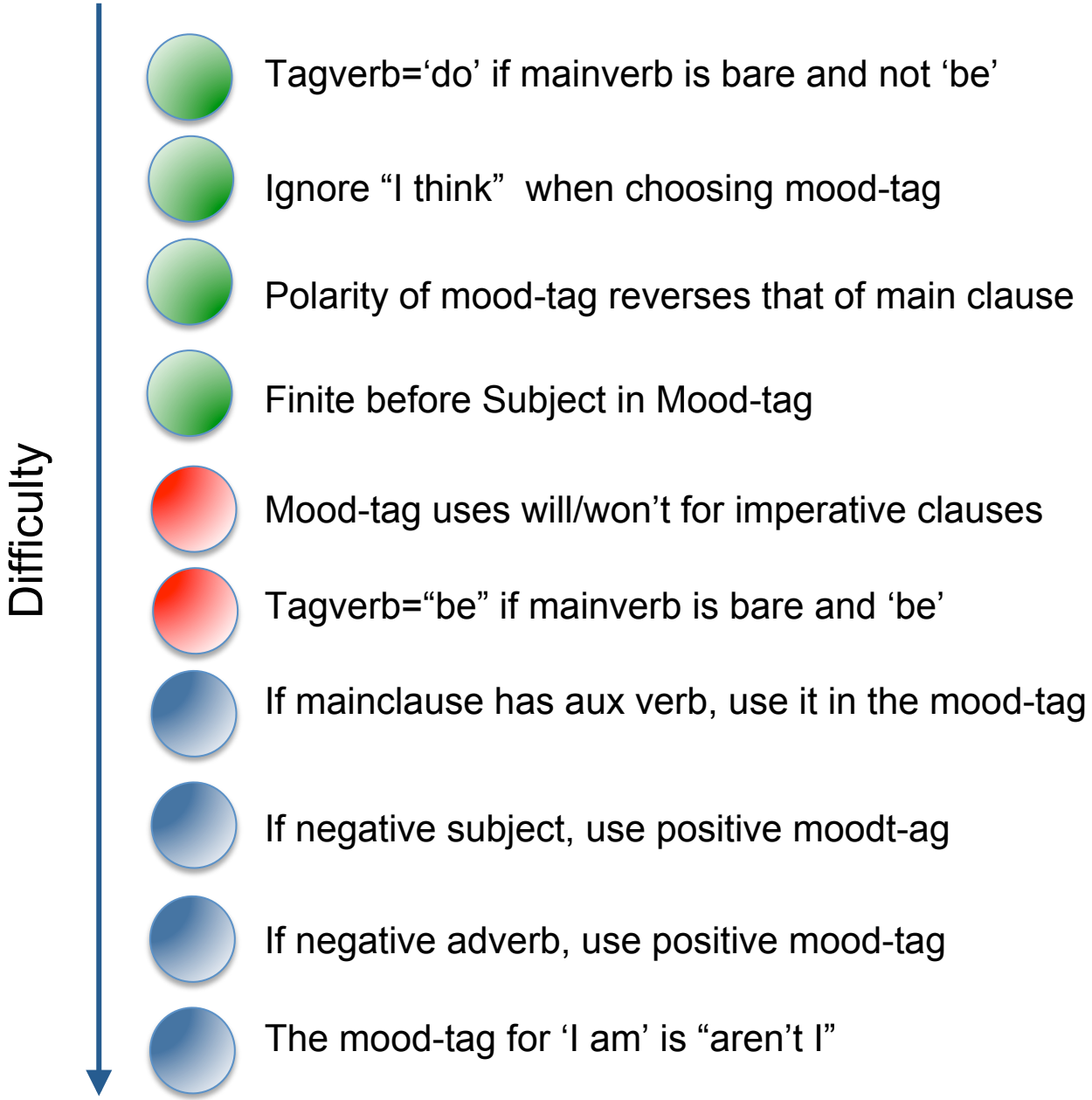
Timeliness: discovering WHEN to teach concepts

Combining multiple sources of evidence

- For each concept-pair, we store two counts:
    1. Cases where A is acquired and B is not
    2. Cases where B is acquired and A is not

- These counts are augmented over whatever quiz data we collect from our students (within the system or from other sources)
- Relative ordering of all concepts recalculated periodically.
- During a student's session, this relative ordering of concepts is used to discover which concept to focus the learner on.

A concrete example: Mood-tag concepts



- Assimiliated Concept
- Partially Assimilated Concept
- Unassimilated Concept

Results from study by Marta Perez

Conclusions

The TREACLE Project

- This talk has presented the work in progress concerning the design of an online blended learning 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:
  - The grammatical concepts that are most critical for our language learners (those that cause the most errors)
    - ordered in terms of overall frequency of observed errors (criticality)
  - Grammatical concepts ordered in difficulty (acquisitional difficulty)

- These are key resources in the adaptive selection of material for the online learner.

The TREACLE Project

- Trail of the system in the next few months in first year Language classes at the Universidad Autónoma de Madrid and the Universitat Politécnica de Valencia.

ADD

- Look up interactive dynamic testing