Towards **targeted** online grammar learning: The Alegro project

Mick O'Donnell (Universidad Autónoma de Madrid)

Penny MacDonald (Universitat Politècnica de València)

The ALEGRO Project

Adaptive
Learning of
English
GRammar
Online

A cooperation between:

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

The Alegro System

- The goal of our project is to develop an **online grammar 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.

The Alegro System

- **Key element**: system is adaptive:

  - it tracks learner's assimilation of concepts via the quizzes and tailors the learner experience on that basis.

Part 1: Adapting to the learner: identifying unlearnt but learnable grammatical content

Language Concepts

By "language concept" we mean:

- A packet of information a speaker needs to produce the language well.

- Mostly equivalent to "rule"

- But can be other, e.g., "class" information

E.g.

○ When referring to two items, use a dual determiner.

○ "both" is a dual determiner.

○ "both"+NOUN is plural

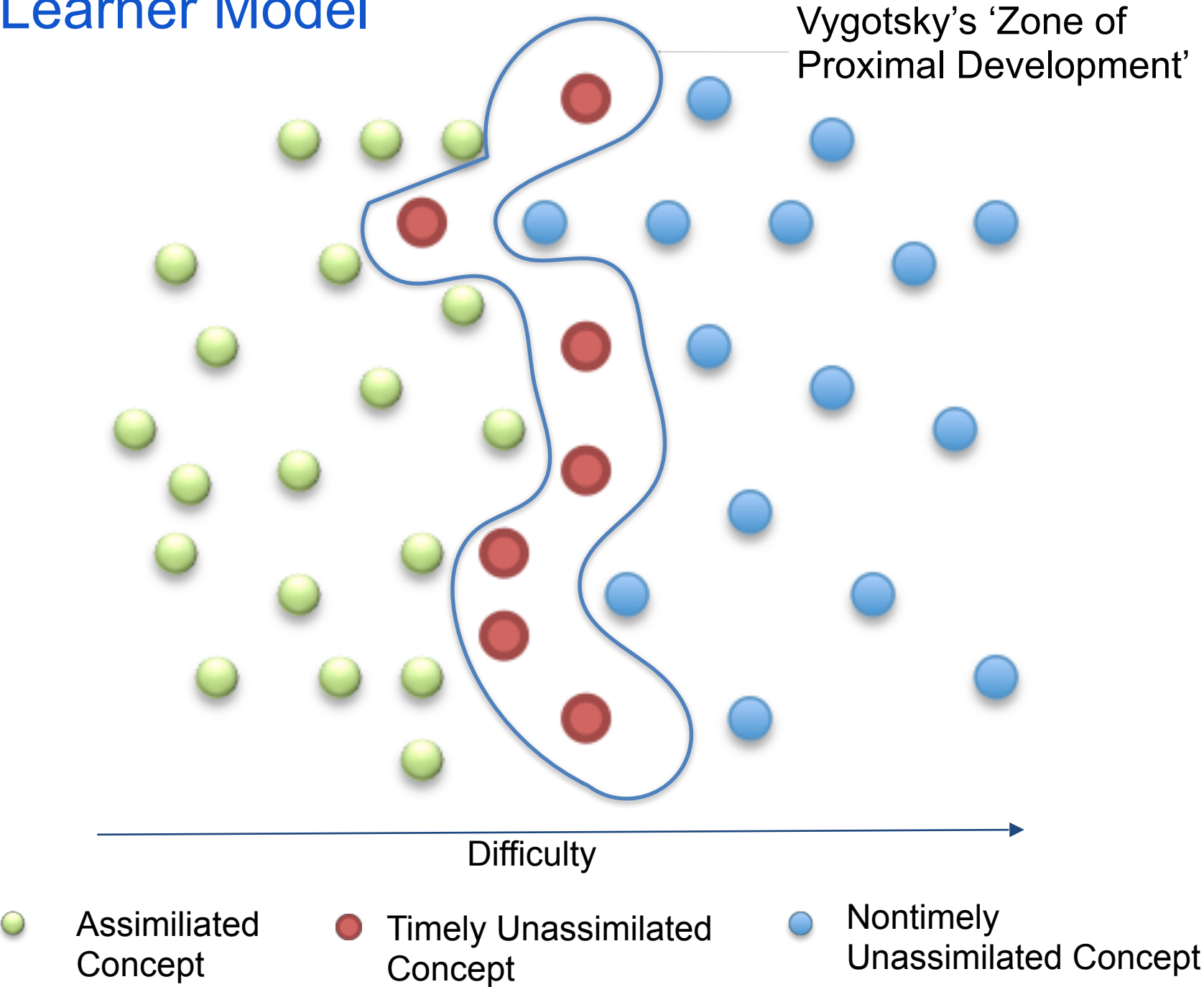# Targeted Learning: focusing

Focus on **learnable content:**

- Learners at a given level of proficiency are not cognitively prepared to learn some more advanced concepts.

- Do not teach what they cannot learn!

Focus on **unlearnt** content

- Quizzes and other means are used to identify the degree to which the learner has assimilated various grammatical concepts.

- Do not teach what they already know!

A Learner Model

Vygotsky's 'Zone of Proximal Development'

Difficulty

| ● Assimiliated Concept | ● Timely Unassimilated Concept | ● Nontimely Unassimilated Concept |

When are grammatical concepts best taught?

- Some theories of language acquisition think that structures are generally learnt at a given level

Not acquired   *Acquired*

A1   A2   B1 │ B2   C1   C2

Grammatical competence with rising proficiency
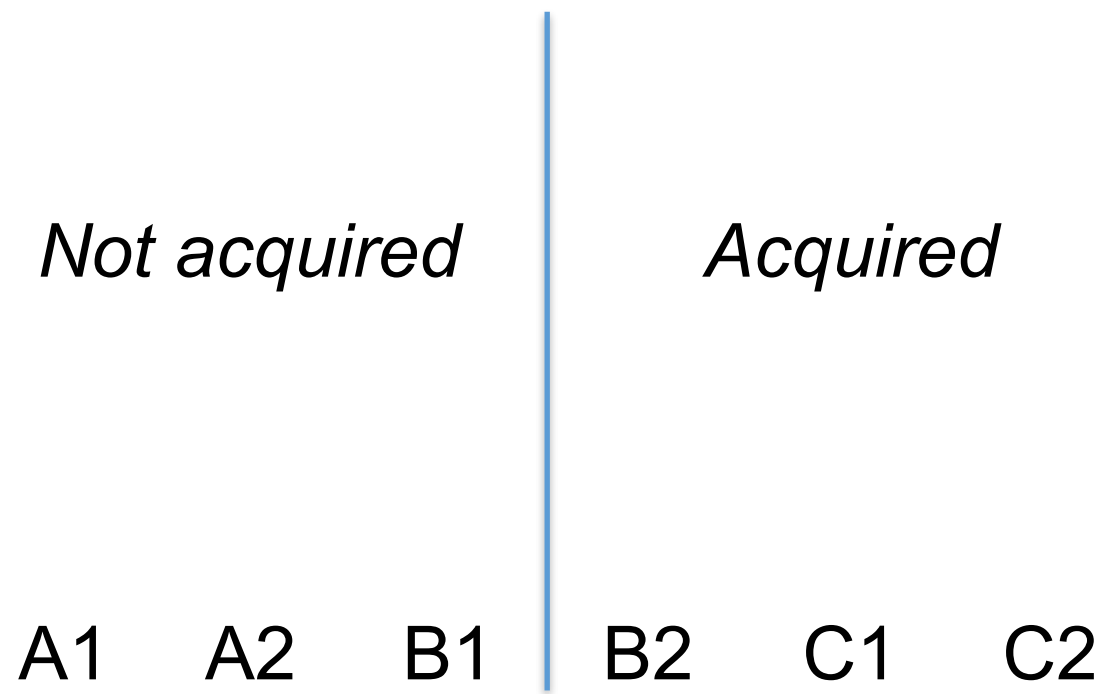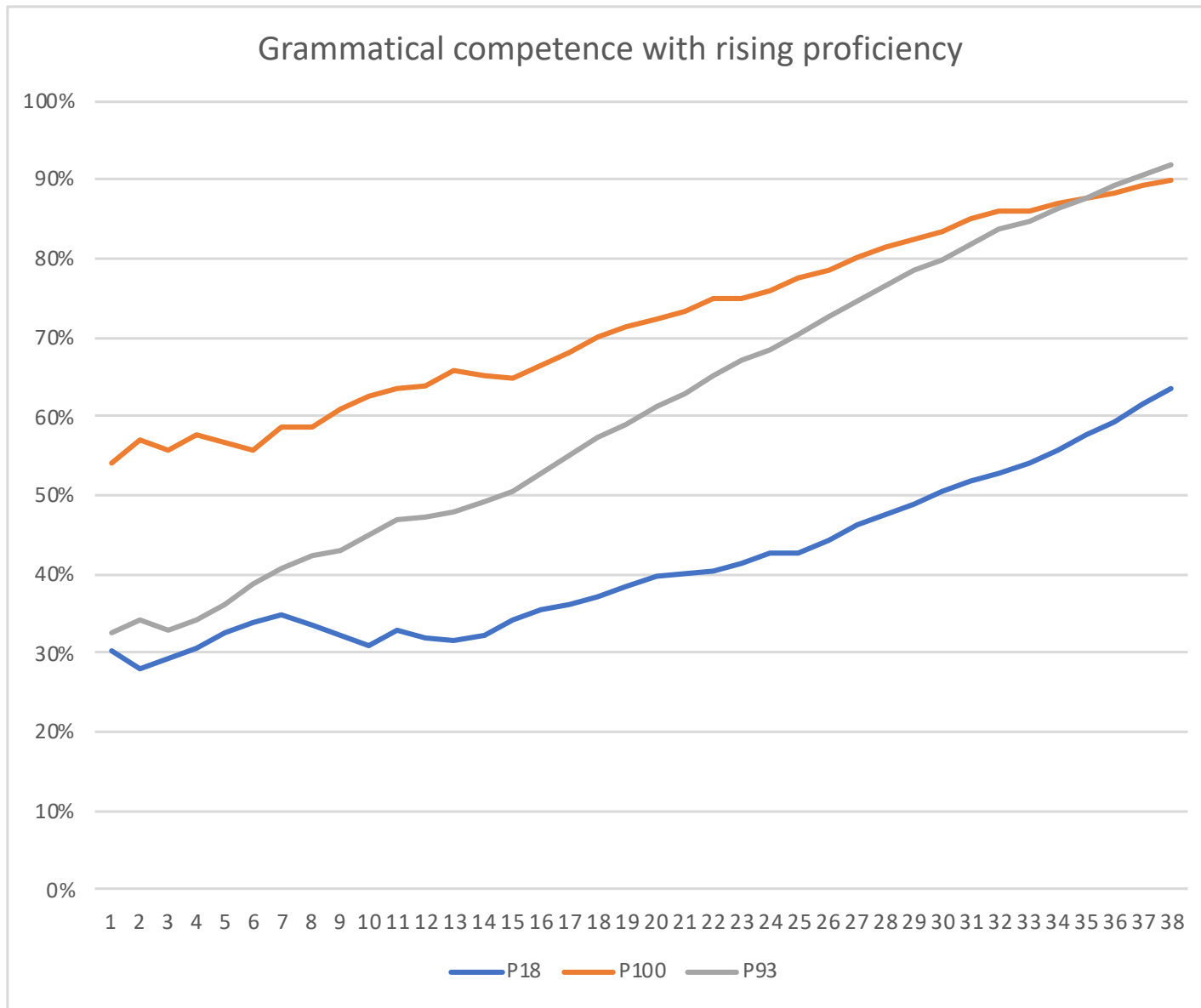
Data collected from 2000 student "prueba de niveles"

At lower levels, random guessing accounts for some correct answers, less with increasing proficiency.

Responses to Multiple choice questions regarding question tags
E.g., "I think it's good [   ]: a) Isn't it? b) Don't I? c) Isn't I?

When are grammatical concept best taught?

- Noticing Hypothesis (Schmidt 1990):

  - Grammatical structures are not acquired until they are noticed.

  - Learner needs to be cognitively prepared to notice the structure.

  - But cognitive preparation is not enough to acquire a structure.

  - The learner needs to encounter the structure, either in language use, or through explicit teaching.
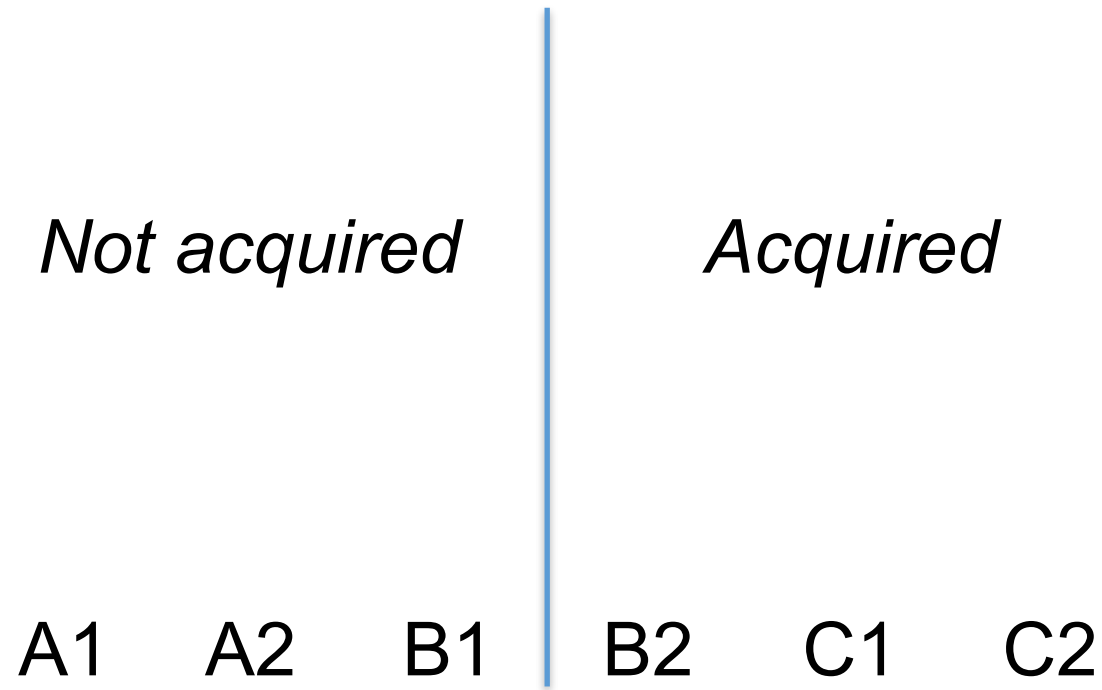
Grammatical competence with rising proficiency
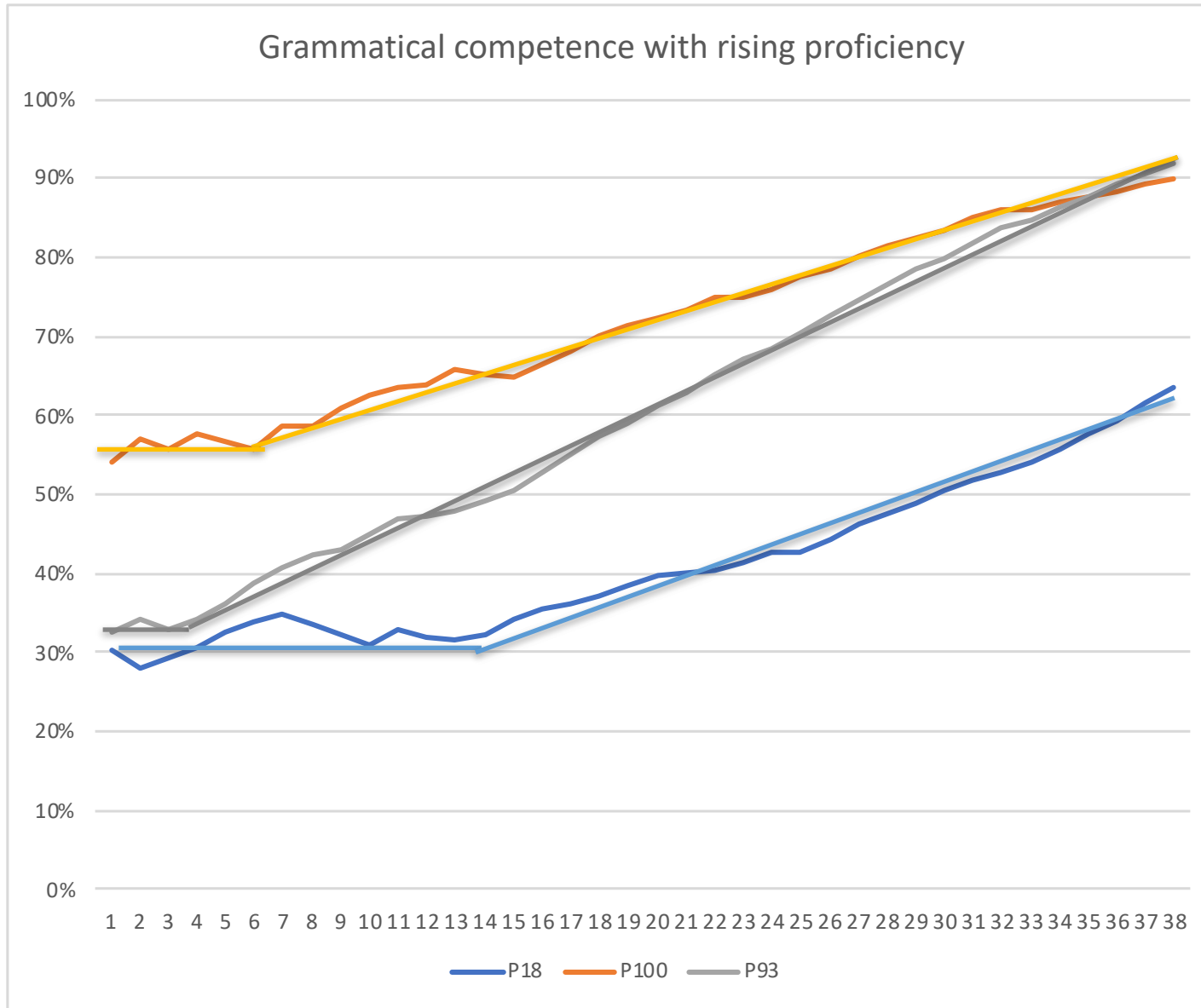
Seems to support the noticing hypothesis:

- At early levels, no acquisition takes place.

- At a certain level, learners start to acquire the feature.

- But learners do not necessarily acquire the feature at that point,

- They require a specific 'noticing' to trigger acquisition.
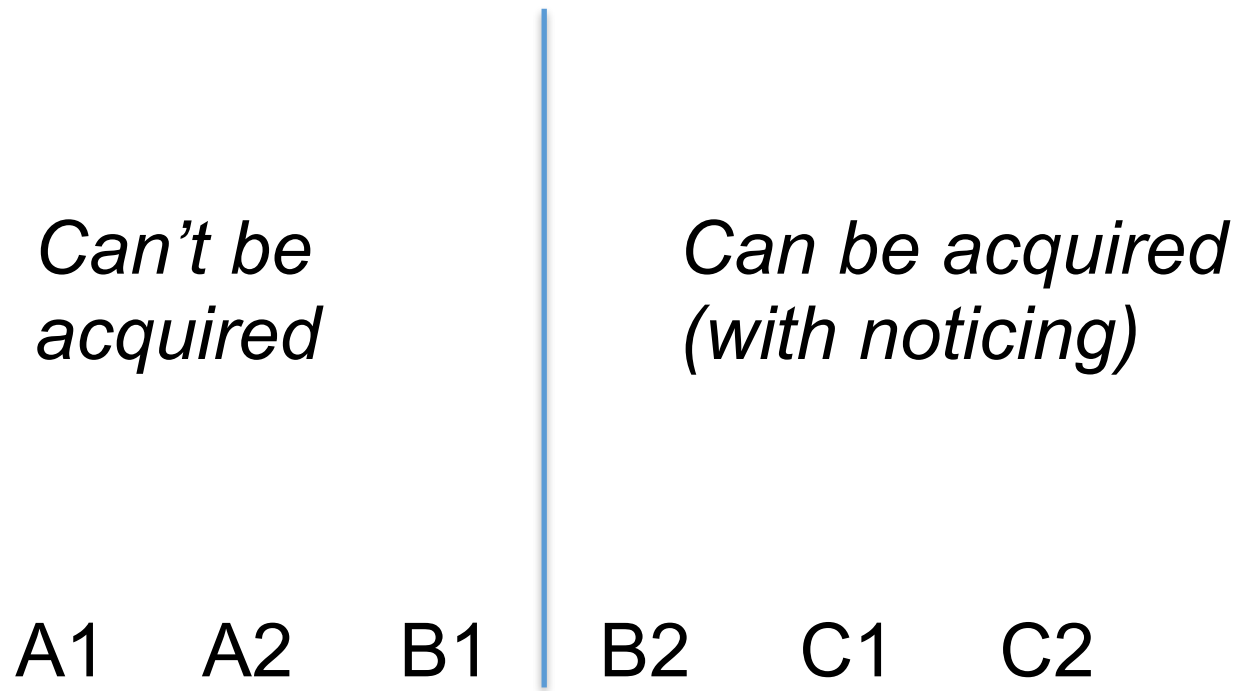
Responses to Multiple choice questions regarding question tags
E.g., "I think it's good [   ]: a) Isn't it? b) Don't I? c) Isn't I?

When are grammatical concept best taught?

Not acquired | Acquired

A1   A2   B1 | B2   C1   C2

When are grammatical concept best taught?

- Noticing hypothesis:

Can't be acquired　　　　*Can be acquired (with noticing)*

A1　　A2　　B1　|　B2　　C1　　C2

Noticing Hypothesis: implications for the system

- When learners are cognitively prepared to acquire a structure, they need a trigger to actually acquire the structure.

- The trigger could be:
  - encountering the structure during reading/ listening in the L2.
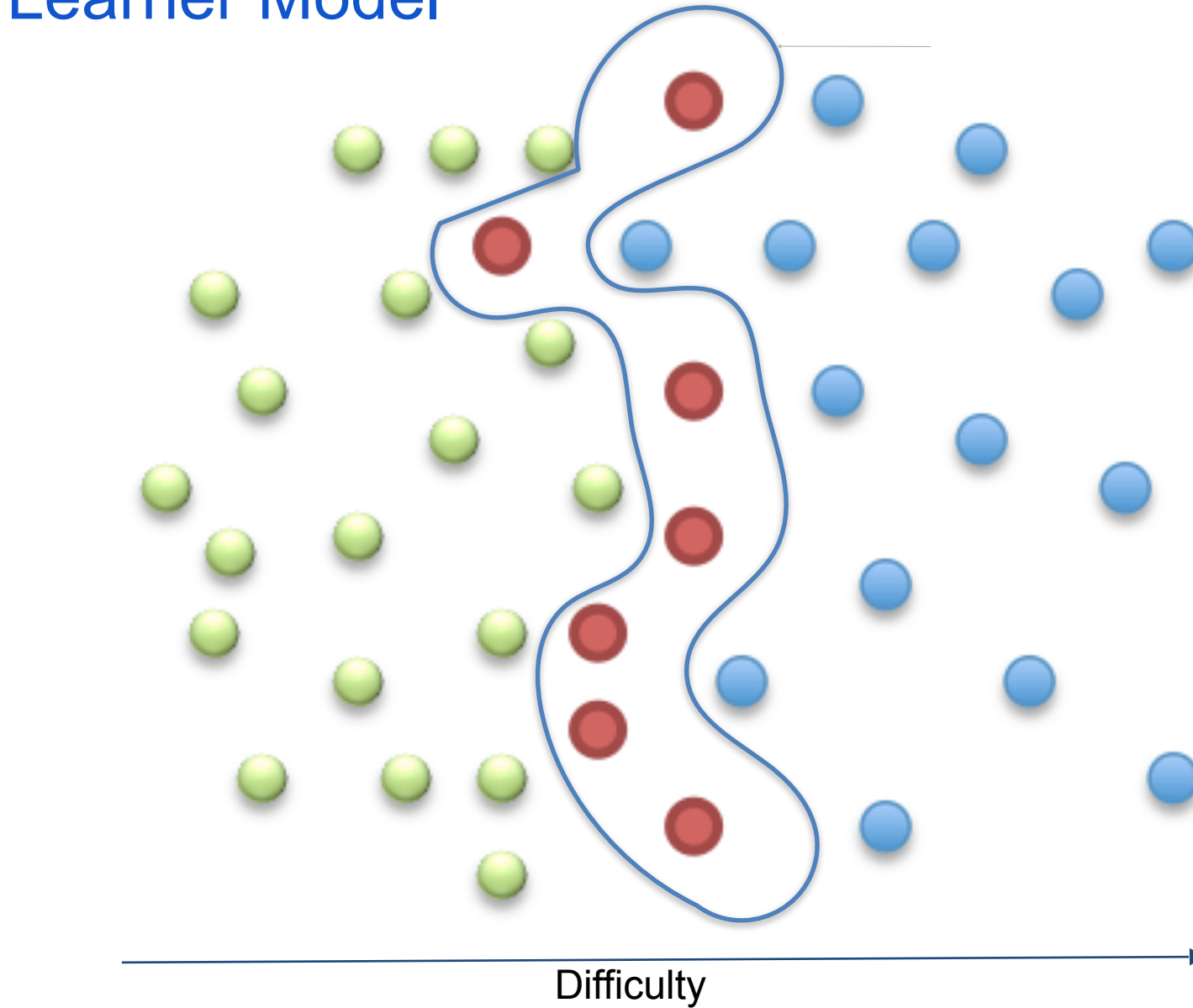 - Explicit teaching of the structure.

Noticing Hypothesis: implications for the system

- Our goal in the system is to identify every grammatical structure the learner is cognitively prepared for, but has not yet acquired.

- THEN, expose them to the structure through explicit explanation of the structure.
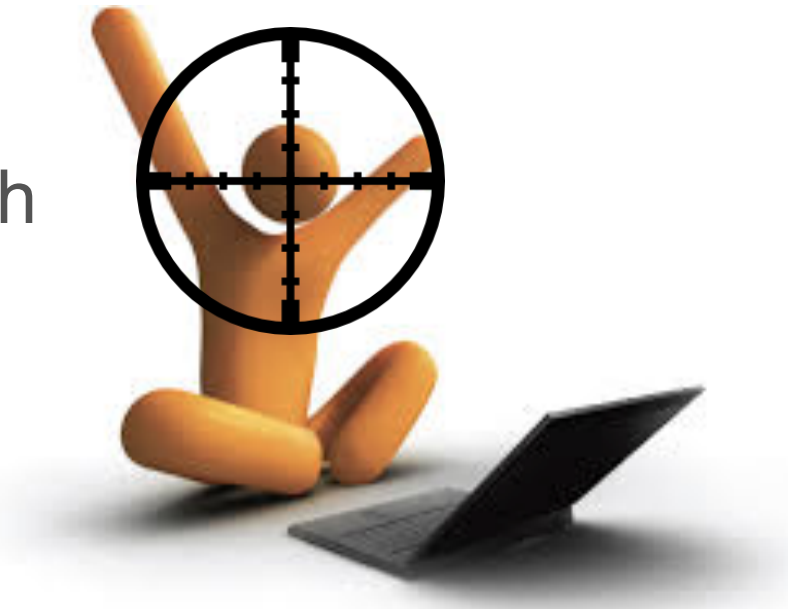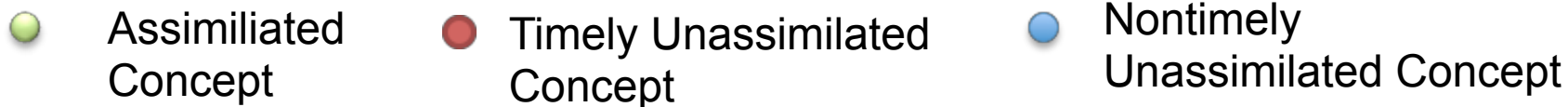
A Learner Model

Focal concepts are those which have not yet been acquired but have the earliest onset of learnability

Difficulty

Assimiliated Concept

Timely Unassimilated Concept

Nontimely Unassimilated Concept

Step 1: Identify general level of difficulty of each grammatical concept

- Responses to quiz questions analysed:
  - **If** more learners have acquired concept A before concept B than vice versa,
    then concept A is acquired earlier than concept B

Step 1: Identify general level of difficulty of each grammatical concept

- Responses to quiz questions analysed:
  - **If** more learners have acquired concept A before concept B, **then** concept A is acquired earlier than concept B

Concept A NOT acquired
Concept B acquired

8%

Concept A NOT acquired
Concept B NOT acquired

47%

Concept A acquired
Concept B acquired

15%

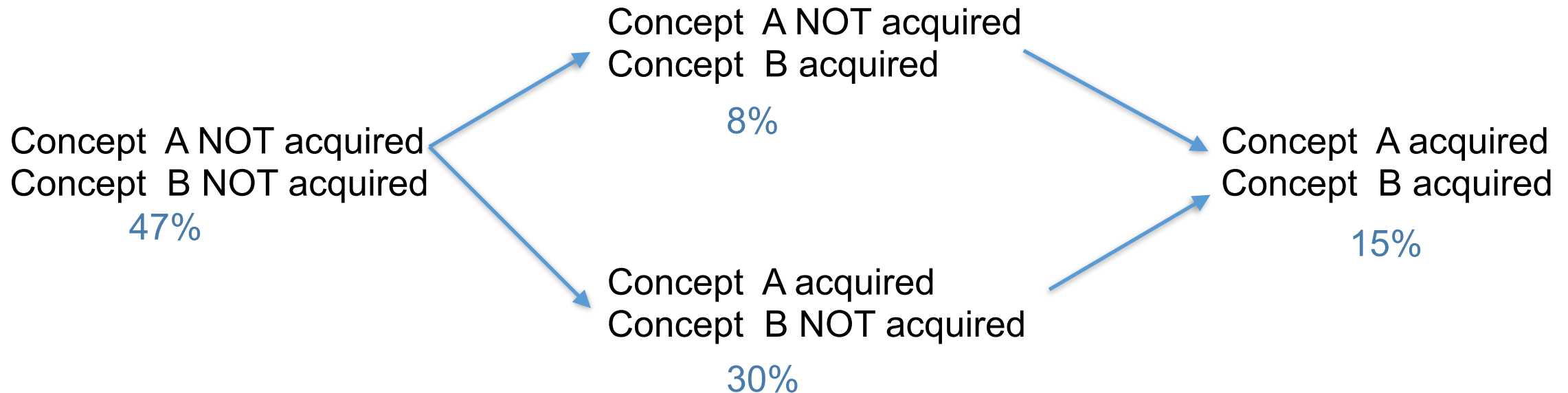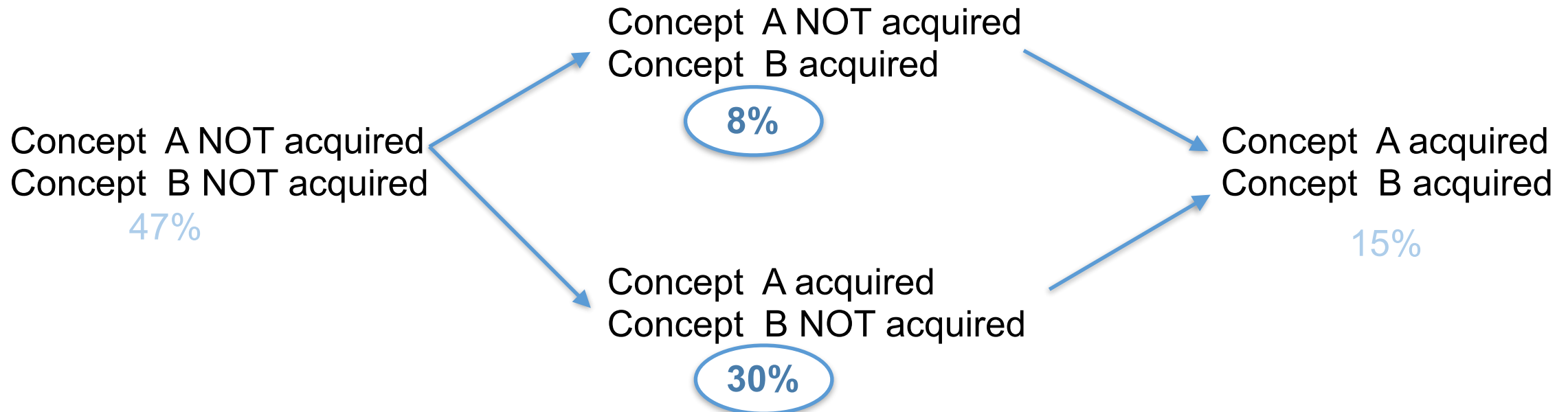Concept A acquired
Concept B NOT acquired

30%

Step 1: Identify general level of difficulty of each grammatical concept

- Responses to quiz questions analysed:
  - **If** more learners have acquired concept A before concept B, **then** concept B is acquired later than concept A

Concept A NOT acquired
Concept B NOT acquired
47%

Concept A NOT acquired
Concept B acquired
8%

Concept A acquired
Concept B NOT acquired
30%

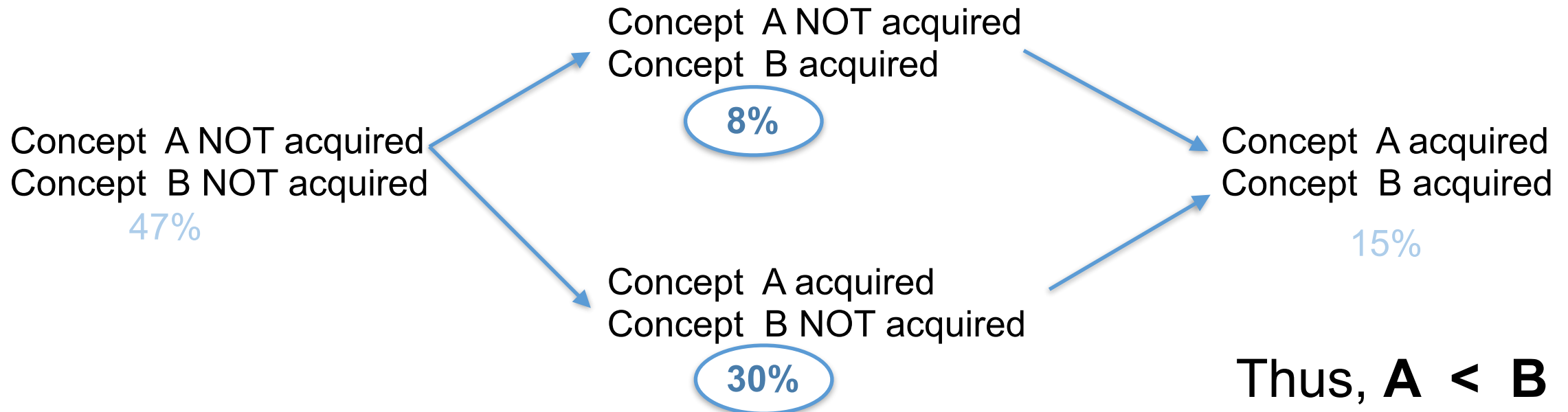Concept A acquired
Concept B acquired
15%

Step 1: Identify general level of difficulty of each grammatical concept

- Responses to quiz questions analysed:
  - **If** more learners have acquired concept A before concept B, **then** concept B is acquired later than concept A
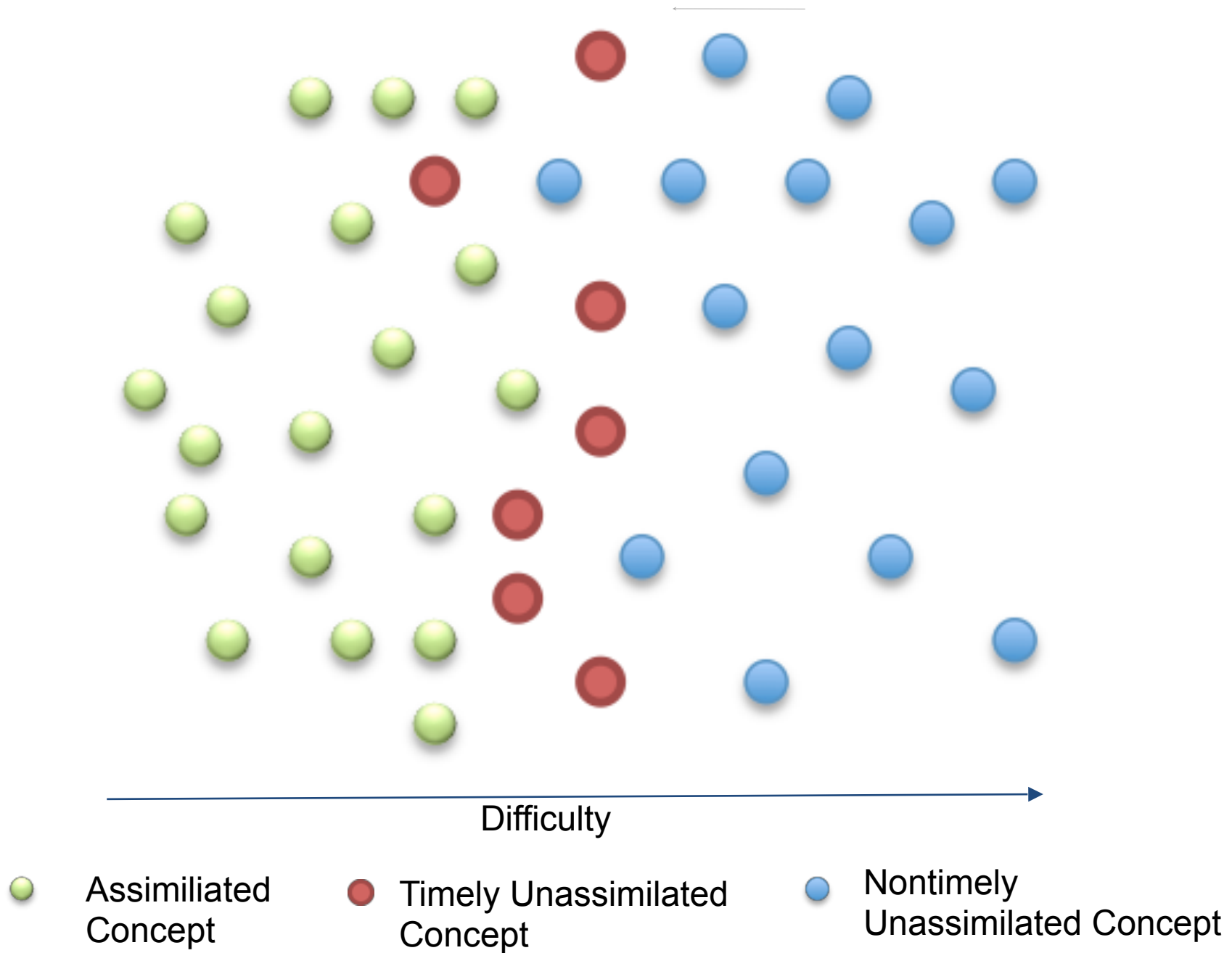
Concept A NOT acquired
Concept B acquired
8%

Concept A NOT acquired
Concept B NOT acquired
47%

Concept A acquired
Concept B acquired
15%

Concept A acquired
Concept B NOT acquired
30%

Thus, **A < B**

We can combine the orderings of pairs of language concepts to produce an overall ordering of learnability of language concepts

A < B

B < C

_____

A < B < C

Difficulty

Assimiliated Concept

Timely Unassimilated Concept

Nontimely Unassimilated Concept

Step 2: Identify an individual learner's level of acquisition of grammatical concepts

- Sentence correctness queries to test if concept is acquired by the learner

Correct or incorrect?

I left my sunglasses inside of the car because the cloudy weather. Correct   Incorrect

- Later: Analysis of learner writing to:
  - identify errors which indicate concept not acquired
  - identify correct applications of a concept

Part 2: Focusing the learner:

Identifying critical language concepts

Critical Language Concepts

- Not all concepts involved in a language are equally important for a given L1/L2 combination:

  - Some can be transferred easily from the L1.

  - Other concepts are difficult to learn but infrequently called upon.

  ➡ Only teach concepts which frequently lead to errors in learner productions.

Critical language concepts

- Our goal: Identify the 1000 language concepts which give most problems to Spanish learners of English.

Noncritical Concepts

- Example of "perro" translated as "dog".
- Concepts:
  - Intralingual concept (ES): "perro" is singular
  - Translation concept: "perro" is translated as "dog"
  - Intralingual concept (EN): "dog" is singular

- No critical concepts here:
  - Spanish natives know that "perro" is singular,
  - They easily learn that "perro" is translated as "dog"
  - They can transfer the singular concept from Spanish to English

Critical Concepts

- Example of "gente" translated as "people".
- Concepts:
  - Intralingual concept (ES): *"gente" is singular*
  - Translation concept: *"gente" is translated as "people"*
  - Intralingual concept (EN): *"people" is plural*

- The fact that the singular number of "gente" is not transferred into English leads to many mistakes by Spanish learners of English.
- The linguistic concept: *"people" is plural* is thus a critical concept for Spanish learners of English.

Critical language concepts

Identifying Critical Language Concepts:

- Our approach: exploration of a learner corpus:

  - Manually tagged 16,000 errors in 110,000 words of learner writing.

  - Identified the 20 most frequent grammatical errors by the learners.

 - Study the grammatical or lexical concepts that lead to these errors.

Corpus

The project makes use of 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

The UPV Learner Corpus

The WriCLE corpus

Annotation Software
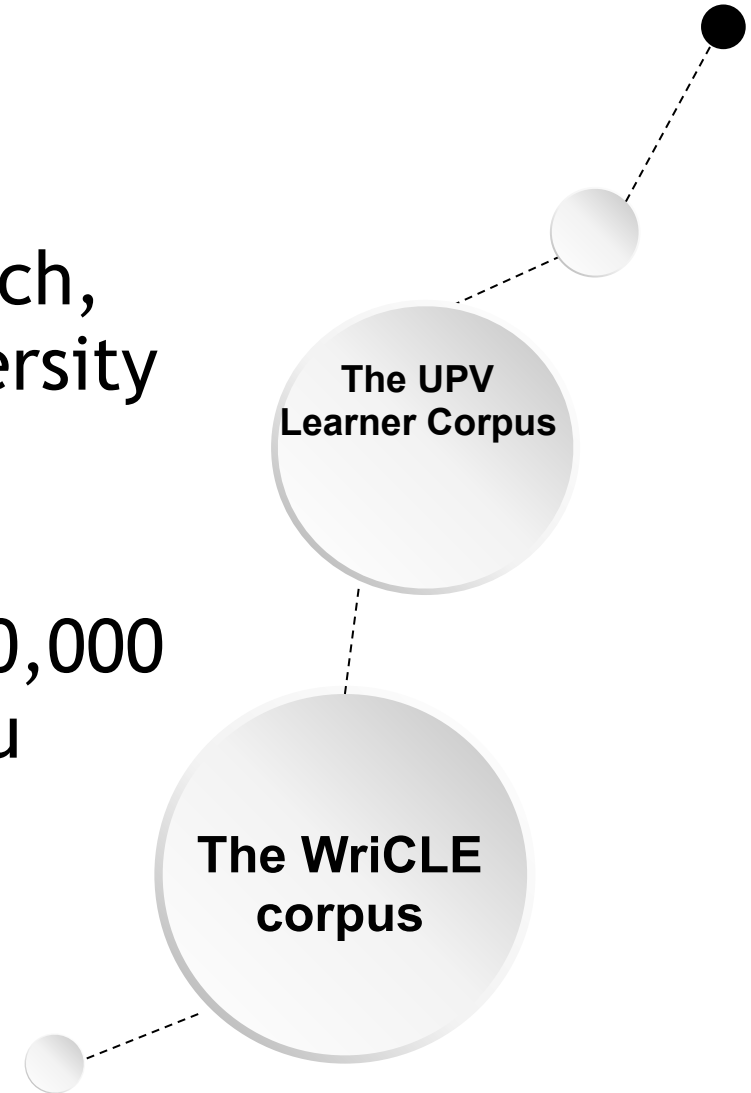
UAM CorpusTool http://www.wagsoft.com/CorpusTool (Free)
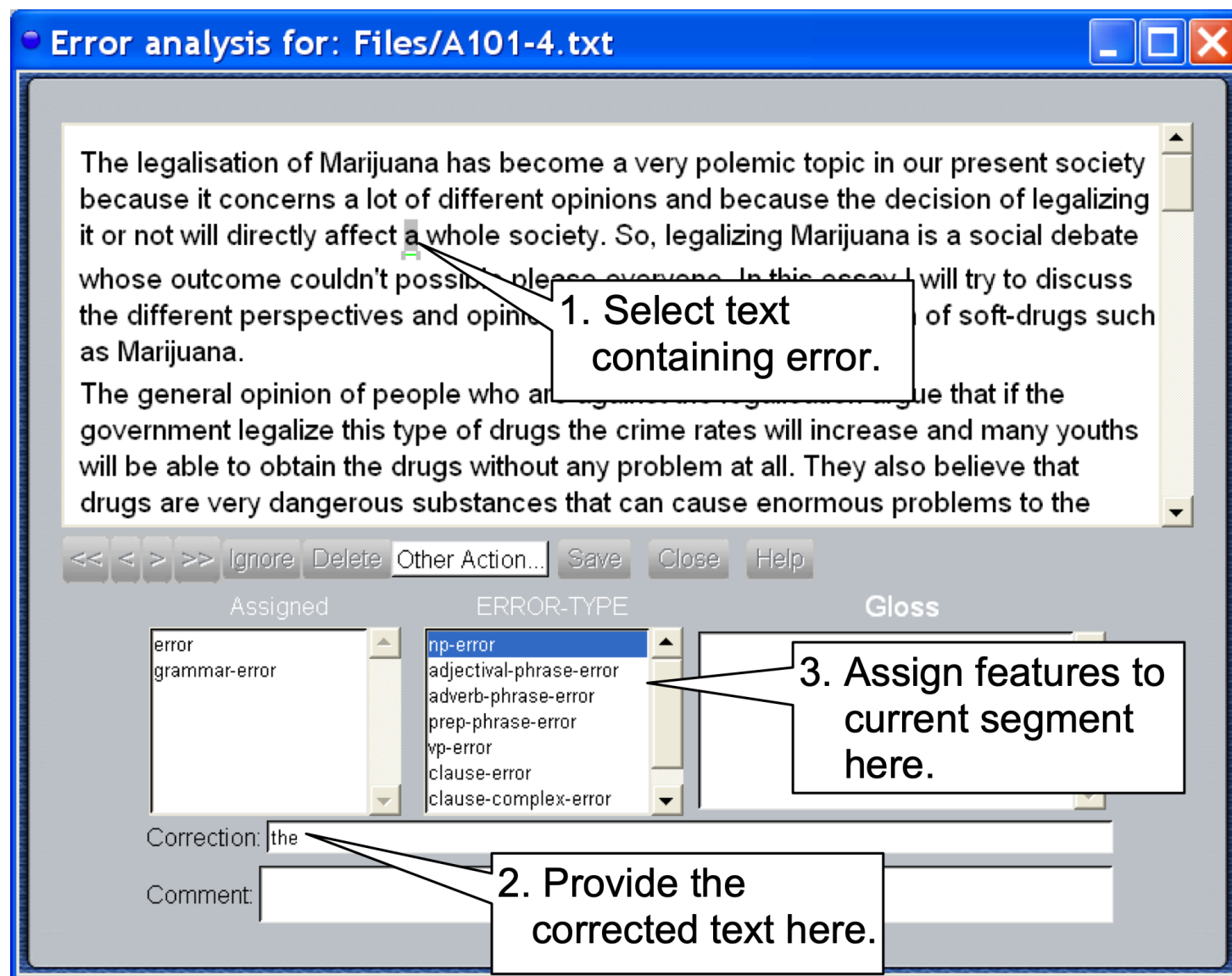
The top 12 most frequent grammatical errors:

Topic	Error	Count	% (of Gramm. Errors)
Determiner	article-wrongly-inserted	1087	**14.7%**
	article-wrongly-absent	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%

MacDonald, P. (2016) "'We al make mistakes!'. Analysing an Error-coded Corpus of Spanish University Students' Written English". Complutense Journal of English Studies 24, 103-129

From grammatical error to language concepts

Towards "teachable" concepts:

- These errors are not at a level that is useful to students:

  - **E.g.,** `article-wrongly-inserted`

  - We cannot teach "use less articles"

  - We need to identify exactly the contexts of use where articles are not needed in English and teach learners these contexts.
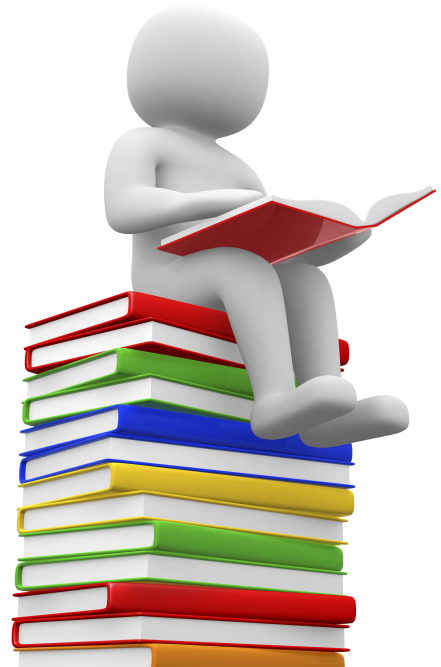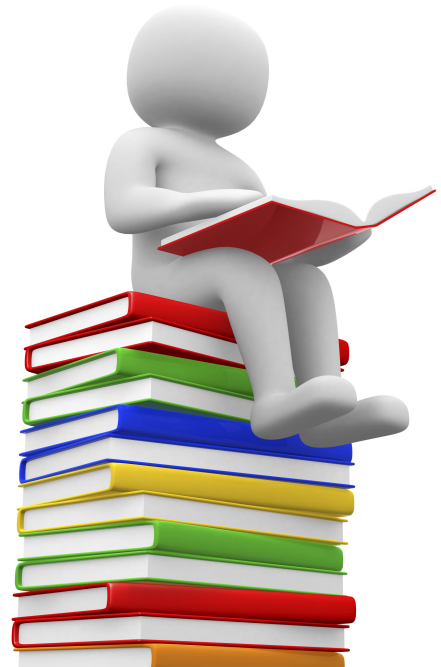
Teachable concept: A rule or concept that the student can apply in their own practice to resolve whether a language instance is valid or not.

Towards teachable concepts

- **Unteachable**: Use less articles!

- **Teachable**: When a plural noun is used to refer to a class of entities (generic reference) with a plural noun, do not use an article.

  ~~The~~ drugs are a problem for some students.

Stage 2: Identifying critical concepts

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"

Results from study by Fiorella Dotti

Stage 2: Identifying critical concepts

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

Stage 2: Identifying critical concepts

- Finer coding of the errors of this category revealed 6 distinct language concepts students need to acquire:

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

Stage 2: Identifying critical concepts

But these errors are critical to different degrees:

Error	Broken concept	% errors
The terrorism is bad.	Generic noncount don't take article	57%
The cats are mammals.	Generic plurals don't take articles	33%
The seventy percent of...	Percentages don't take article	3%
I study in the university	Places of work/internment don't take article	1%
See you after the coffee	Meal-names don't take article	1%
The most of my friends...	*Most* as predeterminer doesn't take article	1%

Stage 2: Identifying critical concepts

All errors involve multiple language concepts

- Even an apparently single-concept error really requires multiple concepts to explain the errors: e.g., subject-finite-agreement

  - Core rule: The Subject and the Finite should agree in number.

- But some cases of errors are explained by other missing concepts:
  - 'people' is plural: *The <u>people is</u> not happy. (*Spanish 'gente' is singular)
  - 'everybody/everyone' is singular.
  - Subject in existential follows verb: *There <u>is many reasons</u>*…
  - 'you' (singular) has plural finite: *<u>You is</u>* …
  - Conjunction of singulars is plural: <u>John and Mary is</u> …

State of the Work

Some topics covered

✓det-absent-required

✓subject-finite-agreement

✓unnecessary-preposition

Others currently in progress

- obligatory-subject-absent
- subject-doubling
- preposition-choice
- det-present-not-required
- Quantification-errors

Others still in progress

Some topics covered
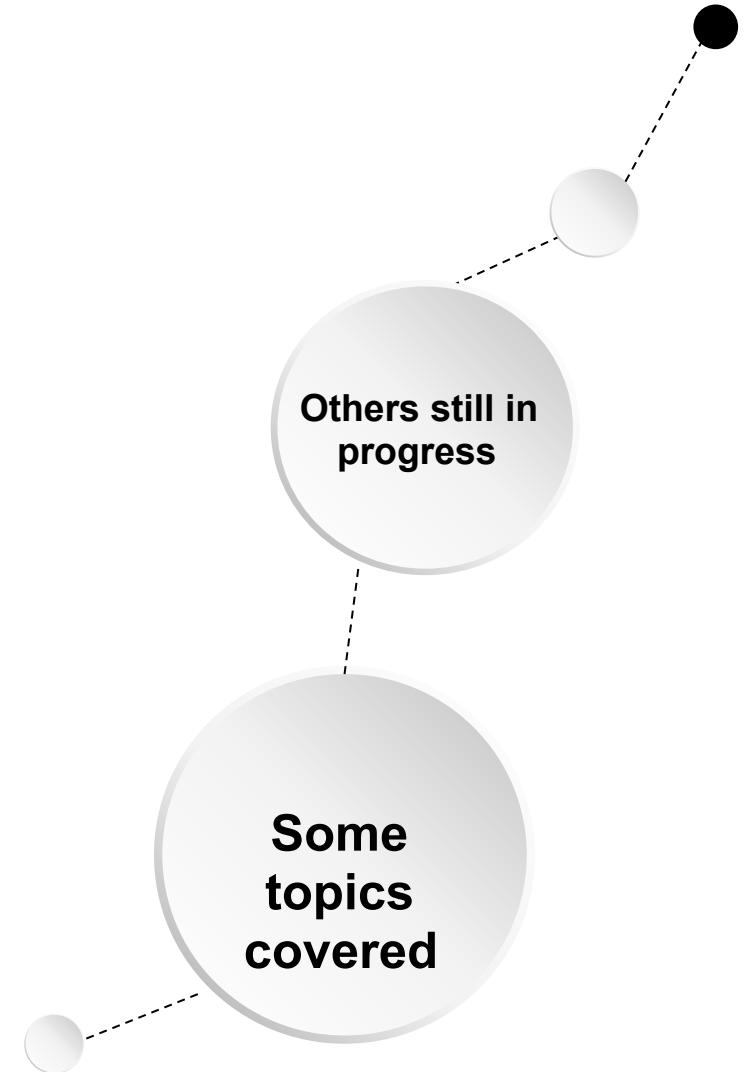
Summary of section

- We have used error annotation of a learner corpus to identify those errors which most affect learner productions.

- Of the most frequent errors, we have explored more deeply to identify the misunderstood language concept which led to the error.

- We then need to add these critical language concepts into our online learning system.

Part 3: Preparing materials for an adaptive learning system

Provide teaching material for critical concepts

- Information is provided to the system in terms of a database of explanations of topics, critical language areas, and critical concepts.

- System delivers this content when contextually appropriate.
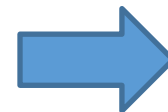
  (system is not a set of pre-written web pages).

Stages of Alegro

- We enter quiz questions to help the learning system diagnose learner competence with respect to these concepts

Part 4: The system in use

Alegro Grammar Learning

Introduction | Review | Study

Welcome to the Alegro system for learning English grammar.

The system is *adaptive* - it identifies your individual grammatical problems and then focuses your work on those problems.

The system offers three interfaces:

- **Review**: allows you to see how much progress you have made on the grammatical concepts. Start here to see which concepts you need to work on.
- **Study**: allows you to review the grammatical concepts covered by the system.
- **Test**: allows you to test how well you know the material you have studied, and identify your problem areas.

Alegro Grammar Learning

This page shows the concepts in your course, and your level of acquisition of these concepts. Study those concepts which have the showing.

Key: **Not Acquired**   **Acquired**   **Unknown**

Area: Prepositions

Module: Preposition-optionality

Topic: Nonverb-related-prep-optionality

Don't use compound prepositions in some contexts   76%   Study

"Most" and currencies do not take a preposition   100%

"Because" must always be followed by the preposition "of" when complemented by a NP   77%   Study

Topic: Verb-related-prep-optionality

Don't use a "to" recipient when placed directly after the verb   7

Screenshot

Areas > Prepositions > Preposition optionality > Nonverb related prep optionality

Concept:

"Because" must always be followed by the preposition "of" when complemented by a NP

When using "because," one must always bear in mind that the preposition "of" needs to come after when complemented by a noun phrase. Some Spanish students, however, omit the preposition and add a whole clause. However, "because of" and "because" are different expressions and, therefore, their construction will differ too.

- WRONG: *I came to the party because **of** my friends insisted.*
- CORRECT: *I came to the party because **of** my friends' insistence.*

The conjunction "because" could also be used, but with a different formulation:

- CORRECT: *I came to the party **because** my friends insisted.*
- WRONG: *I came to the party **because** my friends' insistence.*

Examples:

- *Harry passed the exam because he studied hard.*
- *Harry passed the exam because of his hard work.*

Test me!

(Your acquisition: 77%)

Concept:

"Because" must always be followed by the preposition "of" when complemented by a NP

Test me!

(Your acquisition: 77%)

When using "because," one must always bear in mind that the preposition "of" needs to come after when complemented by a noun phrase. Some Spanish students, however, omit the preposition and add a whole clause. However, "because of" and "because" are different expressions and, therefore, their construction will differ too.

- WRONG: *I came to the party because **of** my friends insisted.*
- CORRECT: *I came to the party because **of** my friends' insistence.*

The conjunction "because" could also be used, but with a different formulation:

- CORRECT: *I came to the party **because** my friends insisted.*
- WRONG: *I came to the party **because** my friends' insistence.*

Examples:

- *Harry passed the exam because he studied hard.*
- *Harry passed the exam because of his hard work.*

Correct or incorrect?

I left my sunglasses inside of the car because the cloudy weather.   Correct   Incorrect

Conclusions

Summary of Talk

- Learners of a language learn best when focused on exactly those concepts:
  - which they have not yet fully acquired
  - which are the least difficult for them to acquire.
- Additionally, learners benefit most from focusing on exactly those concepts which cause most problems to learners of the L1/L2 combination.
- Alegro is intended to focus the learner on exactly these concepts.

Summary of Talk

- We are using a learner corpus to discover exactly which language concepts are most critical for our Spanish learners of English.

- We are using quiz results from many students to place these concepts in order of acquisition difficulty.

- We use quiz results from each individual student to dynamically construct and evolve their learner model (a record of what they have mastered and what they still need to master).

Current Progress

- We have performed one small test of the system with 18 students.

- Around 25% average improvement in concept competence between the pre- and post-test.

- Currently, we are expanding the coverage of the system to a wider range of topics.

- A wide-scale test with over 200 students in the coming weeks.