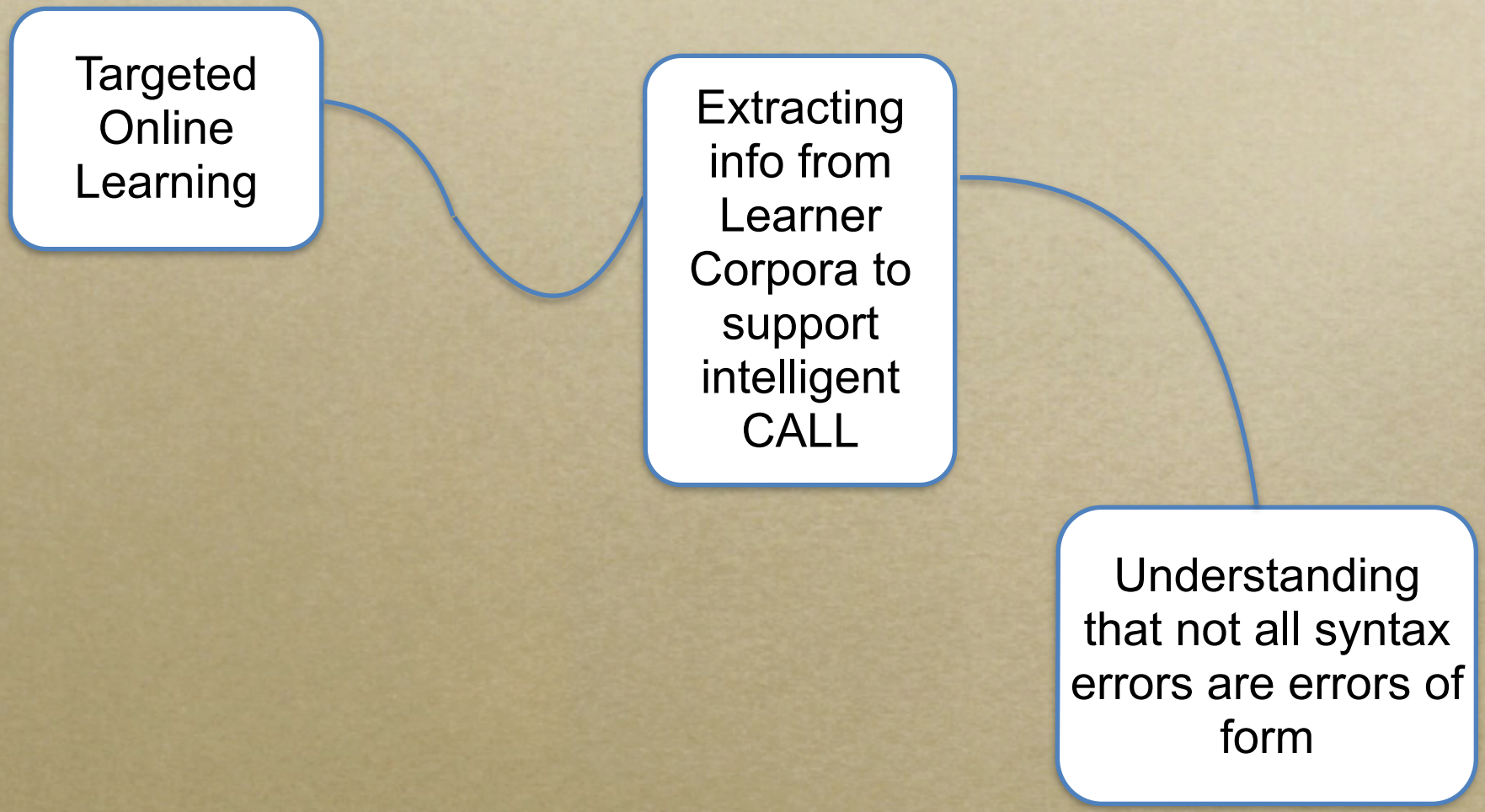
# Exploring learner development in terms of expanding Contexts of Use

Mick O'Donnell
Universidad Autonoma de Madrid

# Overview

- We are working in the context of Spanish University level English education
- We are developing blended learning solutions to English grammar teaching
- We are using learner corpus data to provide the online system models of the language learning process.
- This paper explores some of the issues we faced.

# My Current Research Path



# Part 1:
# A Grammar Learning System that Adapts to the Learner

1.1 Targeted Learning
1.2 Which Concepts to present?
- Critical Concepts
- Timely Concepts

# Introduction: towards targeted Learning

**Starting point**:

- We are building an online system to allow targeted learning of the Grammar of English

**Targeted Learning**: each learner is presented with material exactly suited to their current point of language development

- system tracks which grammatical concepts learner has mastered.
- system presents learner with material just ahead of this point.

# 1000 Critical Concepts

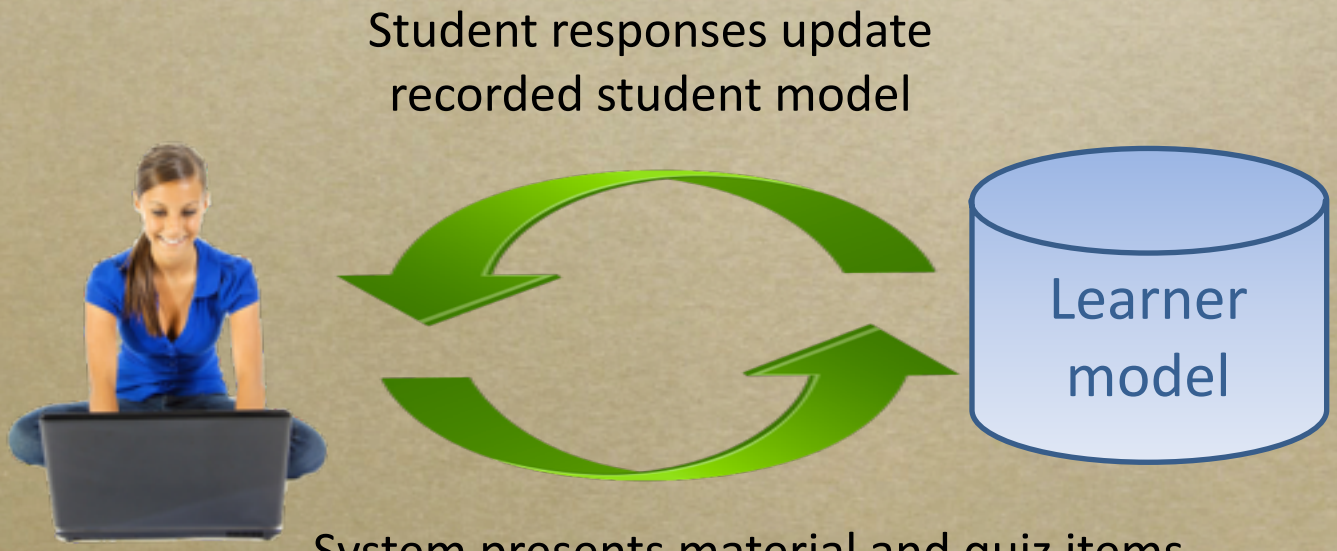
- The system based on idea that there are certain grammatical concepts needed to use a language well.
- System tracks which of these concepts the learner has assimilated, and which are still to assimilate.
- All material in the system (teaching material, quiz items, etc.) indexed in terms of these concepts.
- So, material to present next to the learner is selected in terms of which of of these concepts most valuable to the learner at this point of time,

# Grammatical Concepts for use of “much”

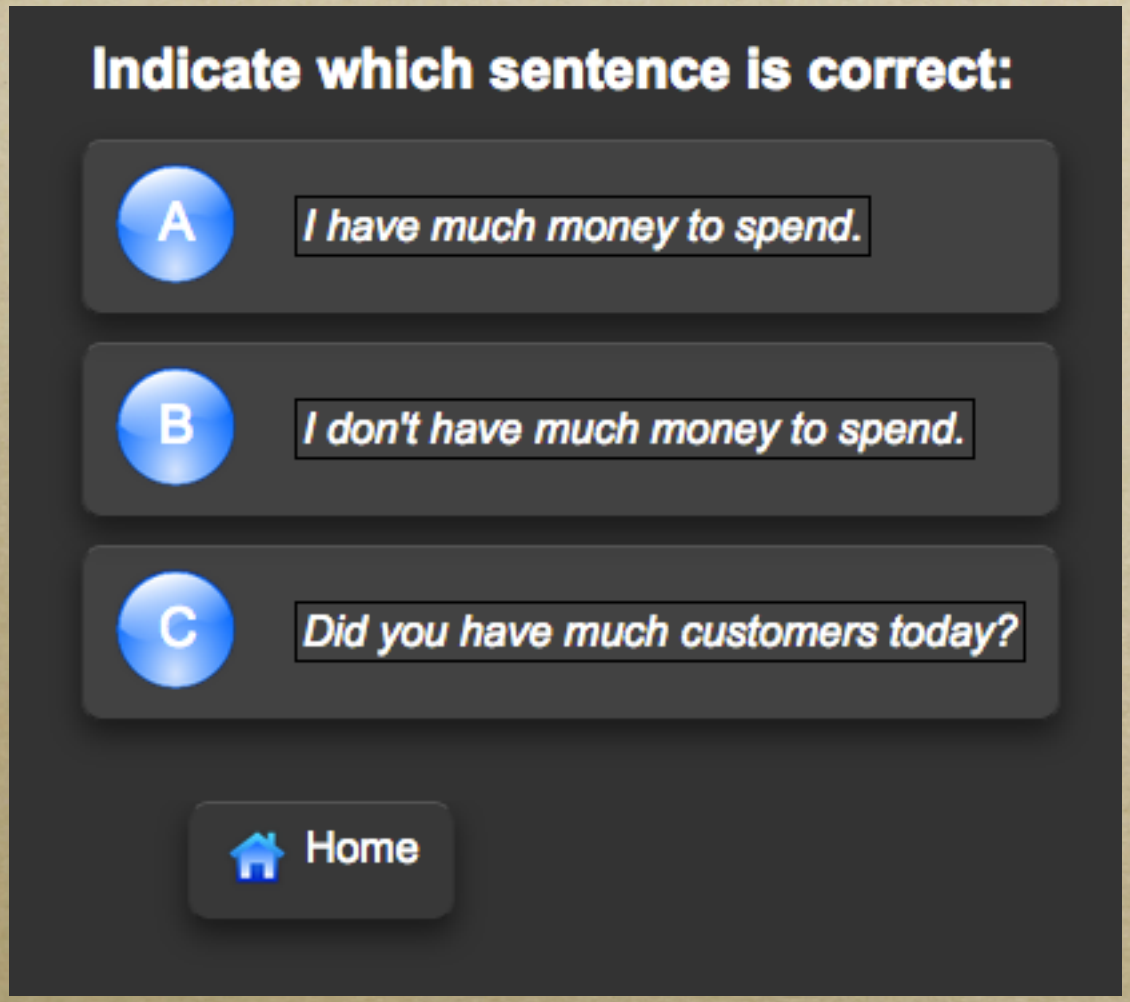
1. ‘much’ is used with mass nouns only.
   *much water*
   *much apples*
2. ‘much’ is not usually used in affirmative sentences.
   *I have much water*
3. ‘much’ can be used in negated statements.
   *I don’t have much water*
4. ‘much’ can be used in a positive clause embedded in a negative one.
   *I don’t think that we have much water.*
5. ‘much’ can be used in questions.
   *Do you have much water?*
6. ‘much’ can be used in affirmative sentences when it is preceded by "so", "too" or "as" (intensifiers/comparatives).
   *I have too much time.*
   *He has so much hair.*
   *You have as much money as me.*

# Modelling the LEARNER

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



# A Sample Quiz Question



# The Question Database

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

| ***Answer*** | ***Concepts Broken*** | ***Concepts Confirmed*** |
|---|---|---|
| *I have much money to spend.* | *much-not-in-simple-decl* | *much-with-mass* |
| *I don't have much money to spend.* | | *much-with-neg; much-with-mass* |
| *Did you have much customers today?* | *much-with-mass* | *much-with-question* |

# Part 1:
# A Grammar Learning System that Adapts to the Learner

1.1 Targeted Learning

**1.2 Which Concepts to present?**

- Critical Concepts
- Timely Concepts

# Towards Targeted Learning I
## Material selection

Two aspects in judging importance of grammatical concepts for a learner:

1. Criticality: Language concepts that are real and observable problems to language learners as a whole (of a particular L1)

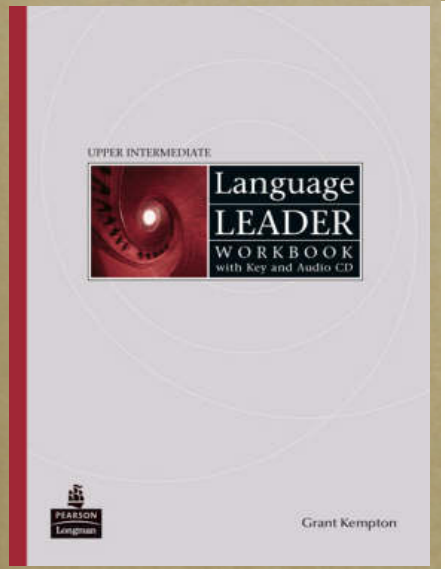
2. Timeliness: Language concepts which are critical to the particular learner at this point of time.

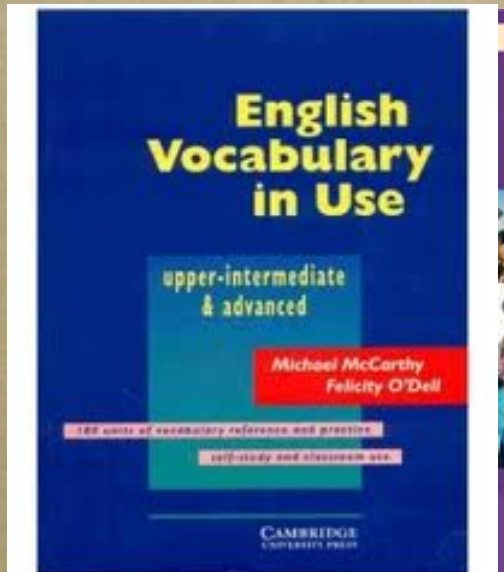
# Criticality

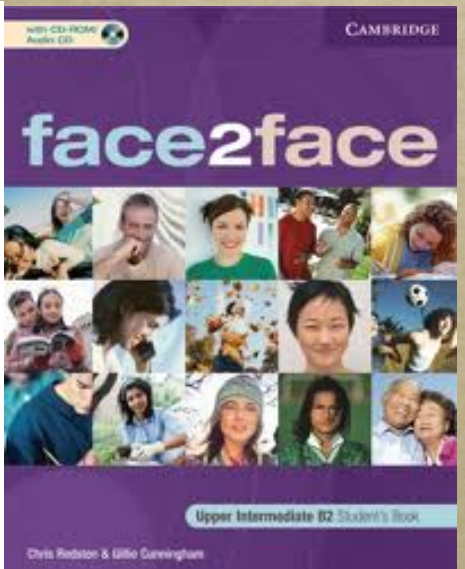
- Learning a foreign language requires mastering thousands of grammatical concepts.
- But many of these concepts are easily acquired, directly transferable from the mother tongue, e.g., English and Spanish share passive structure, progressive aspect, etc.
- We should not waste the time of our learners teaching structures they can transfer from their mother tongue.
- **Critical concepts**: exactly those linguistic concepts that demonstrably cause problems for learners from a particular L1.

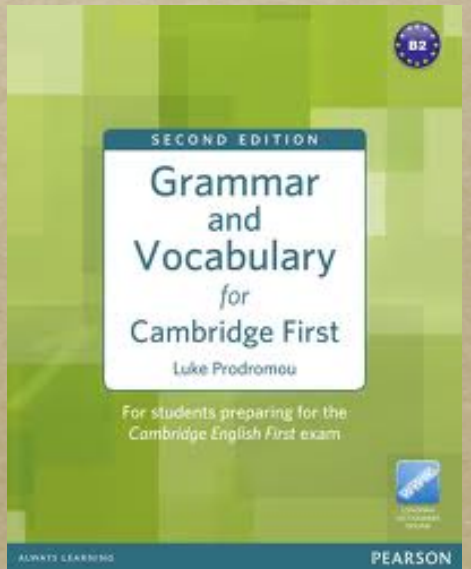
# L1-Specific Criticality

- Nearly all English resources are aimed at a mythical “average-L1” learner.
- The linguistic concepts which are most critical for one mother tongue may not be critical for another.



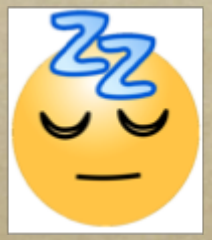



# Timely Concepts

Key concept: Vygotsky's ***Zone of proximal Development*** (or the Goldilocks principle)

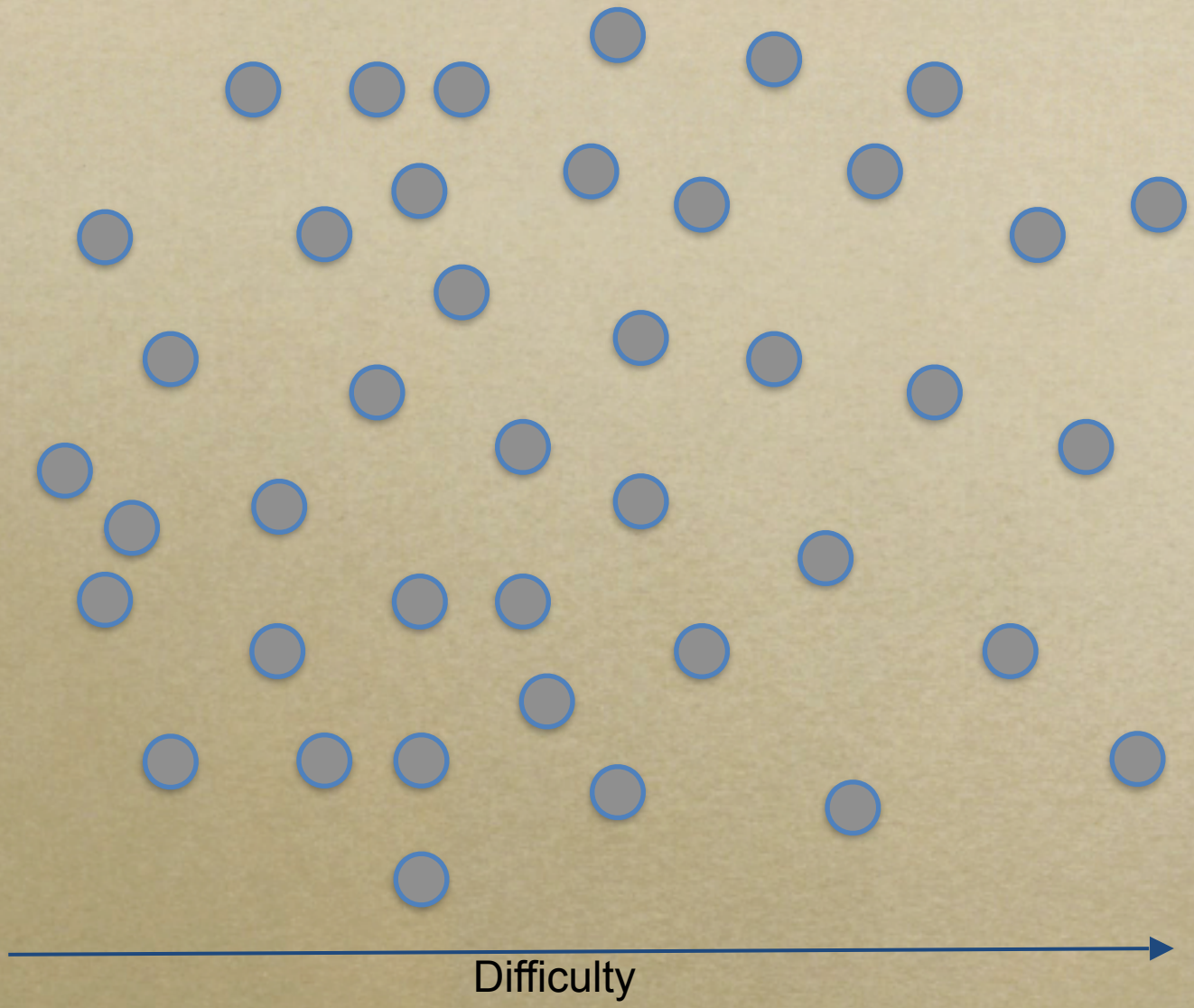


Learners are bored by material which they already know.

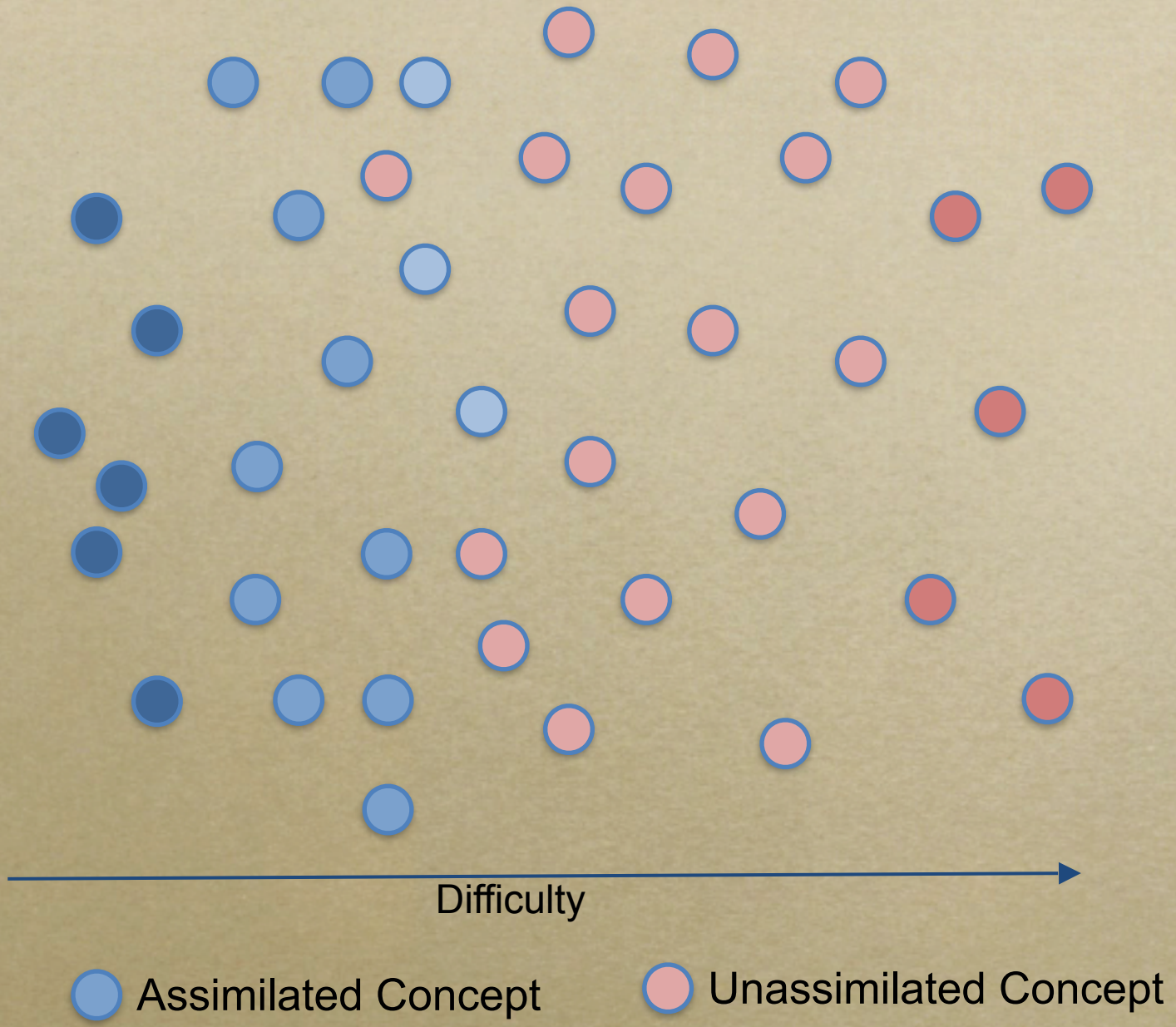
Learners find it difficult to assimilate material which is too advanced for them (outside of our current cognitive ability)

We learn best when presented with material just at our current level (Engagement is maximised)

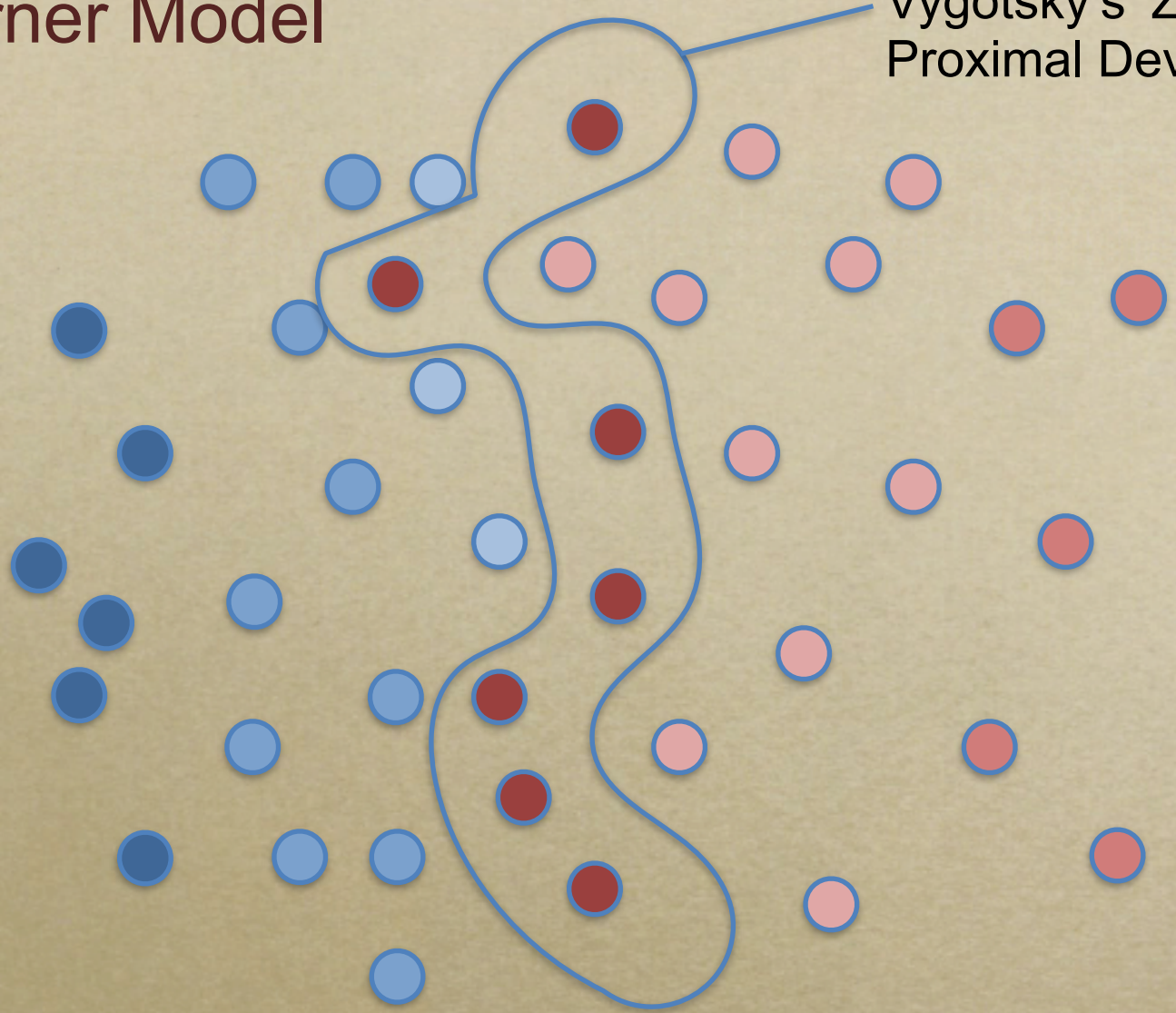
# A Learner Model



# A Learner Model



# A Learner Model




Assimilated Concept

Timely Unassimilated Concept

Untimely Unassimilated Concept

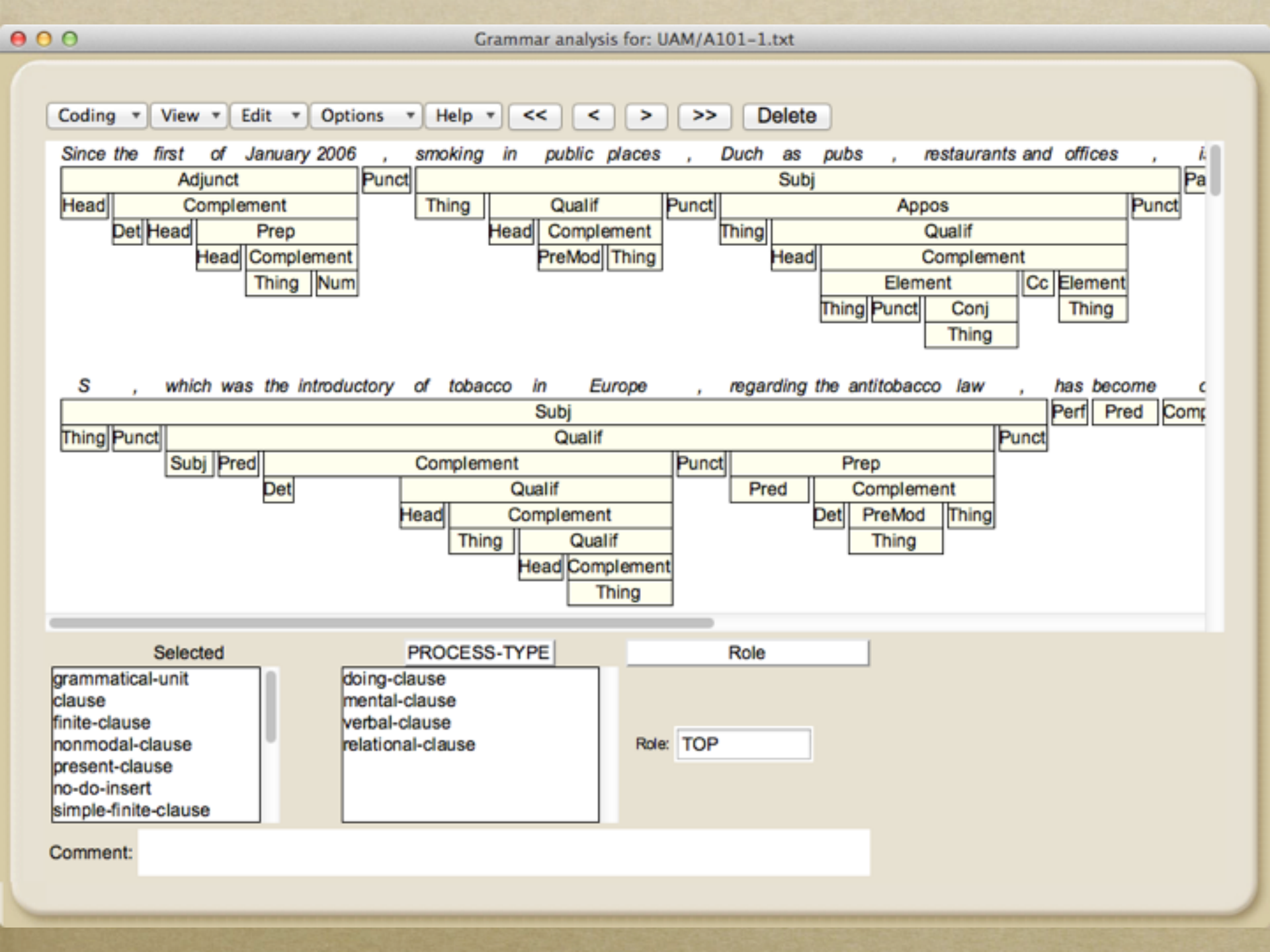
# Part 2:
# Using Learner Corpora To Discover Concept Criticality and Difficulty

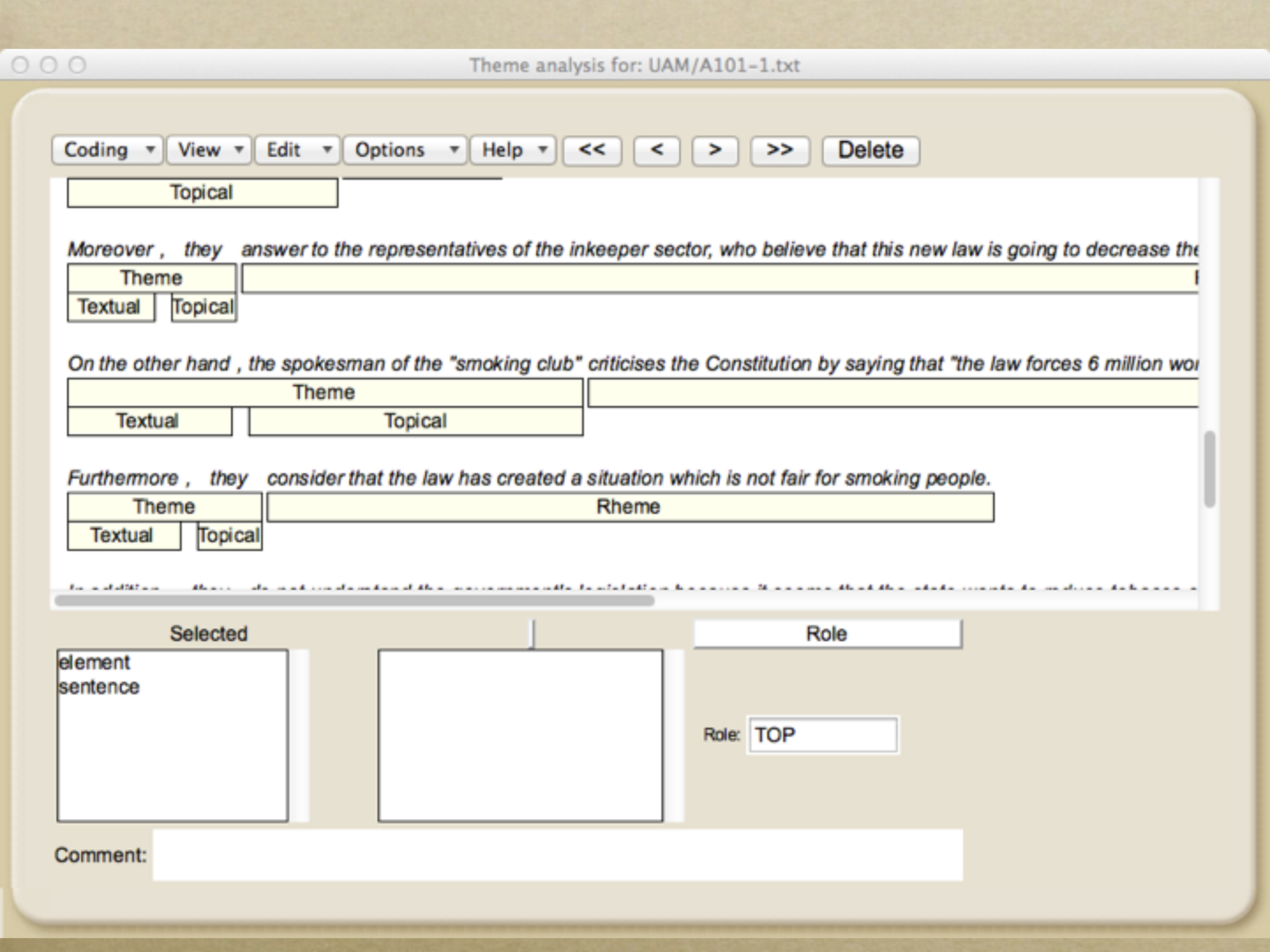
# Modelling Learners through analysing their output

- The online learning system needs to contain ‘representations’ of typical language learners.
- These representations should be based on the particular L1 that we are teaching.
- The representations should take into account the evolving nature of abilities as the learner advances in proficiency.
- For this purpose, we collect written productions of these learners (a learner corpora) and explore their abilities and lacks at each stage.

# Modelling Learners through their production

- The Treacle project uses a corpus of Learner English produced by Spanish University learners:
  - **WriCLE corpus**: 500,000 words (521 essays) collected by Paul Rollinson at UAM (1st year and 3rd year of English Studies)
  - **UPV Learner Corpus:** 150,000 words of shorter texts by ESP students at Universidad Politecnica de Valencia.


Grammar analysis for: UAM/A101-1.txt
Coding
View
Edit
Options
Help
<<
<
>
>>
Delete
Since the first of January 2006 , smoking in public places , Duch as pubs , restaurants and offices , i
Adjunct
Punct
Subj
Pa
Head
Complement
Thing
Qualif
Punct
Appos
Punct
Det
Head
Prep
Head
Complement
PreMod
Thing
Thing
Qualif
Head
Complement
Thing
Num
Element
Cc
Element
Thing
Punct
Conj
Thing
Thing
S , which was the introductory of tobacco in Europe , regarding the antitobacco law , has become c
Subj
Perf
Pred
Comp
Thing
Punct
Qualif
Punct
Subj
Pred
Complement
Punct
Prep
Det
Qualif
Pred
Complement
Head
Complement
Det
PreMod
Thing
Thing
Qualif
Thing
Head
Complement
Thing
Selected
PROCESS-TYPE
Role
grammatical-unit
clause
finite-clause
nonmodal-clause
present-clause
no-do-insert
simple-finite-clause
doing-clause
mental-clause
verbal-clause
relational-clause
Role:
TOP
Comment:


Theme analysis for: UAM/A101-1.txt
Coding
View
Edit
Options
Help
<<
<
>
>>
Delete
Topical
Moreover , they answer to the representatives of the inkeeper sector, who believe that this new law is going to decrease the
Theme
Textual
Topical
On the other hand , the spokesman of the "smoking club" criticises the Constitution by saying that "the law forces 6 million wo
Theme
Textual
Topical
Furthermore , they consider that the law has created a situation which is not fair for smoking people.
Theme
Rheme
Textual
Topical
Selected
element
sentence
Role
Role: TOP
Comment:

Coding | View | Edit | Options | Help | << | < | > | >> | Delete

*In this essay , I intend to present different points of view about the new antitobacco law .*

| Circumstance | Senser | Process | Phenomenon | | |
|---|---|---|---|---|---|
| | | | Process | Goal | Circumstance |

*This law establishes smoking zones in pubs, restaurants etc .*

| Actor | Process | Goal |
|---|---|---|

*It limits publicity refering to tobacco and hardens the normative of smoking in public places .*

| Element | | | | Element | | |
|---|---|---|---|---|---|---|
| Actor | Process | Goal | Circumstance | Process | Goal | Circumstance |

*In addition , it attempts to improve spanish citizens health , as it is a fact that the first cause of death in our c*

| Circumstance | Actor | Process | Goal | | Circumstance | | | |
|---|---|---|---|---|---|---|---|---|
| | | | Process | Goal | Process | Attribute | Carrier | |
| | | | | | | | | Carrier |

Selected

ideational-unit
circumstance

Role

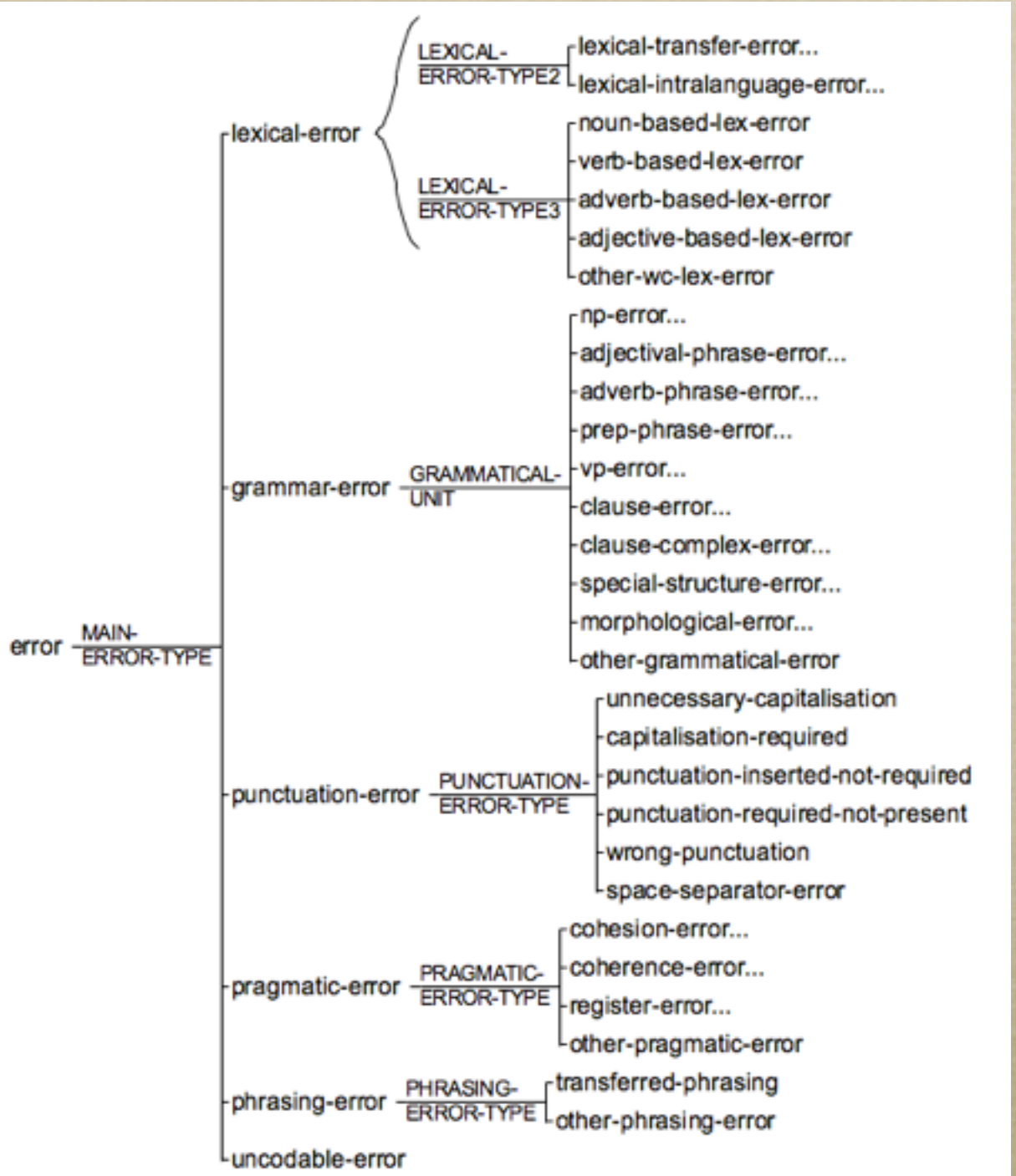
Role: Circumstance

Comment:

# Error-annotation

- 110 student essays across 6 proficiency levels were manually annotated for errors.
- 116,000 words
- 16,600 errors identified.

- Reference:  Murcia Bielsa & MacDonald, 2013


error
MAIN-ERROR-TYPE
lexical-error
LEXICAL-ERROR-TYPE2
lexical-transfer-error...
lexical-intralanguage-error...
LEXICAL-ERROR-TYPE3
noun-based-lex-error
verb-based-lex-error
adverb-based-lex-error
adjective-based-lex-error
other-wc-lex-error
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...
morphological-error...
other-grammatical-error
punctuation-error
PUNCTUATION-ERROR-TYPE
unnecessary-capitalisation
capitalisation-required
punctuation-inserted-not-required
punctuation-required-not-present
wrong-punctuation
space-separator-error
pragmatic-error
PRAGMATIC-ERROR-TYPE
cohesion-error...
coherence-error...
register-error...
other-pragmatic-error
phrasing-error
PHRASING-ERROR-TYPE
transferred-phrasing
other-phrasing-error
uncodable-error

# Identifying Critical Concepts

- Observe where learners of a particular L1 go wrong:
  - **Error Analysis** to identify which linguistic structures/words cause errors most frequently.
  - **Syntactic analysis** to identify linguistic items which the learner is avoiding or over-using

By focusing our teaching effort on those structures which give learners most problems, learning time is more productively used.

# Identifying Critical Concepts

## Critical Grammar Concepts in terms of errors for Spanish learners of English:

**Grammatical Errors in order of Frequency in Treacle Error Corpus:**

- 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
- ...

# Calculating Concept **Difficulty**

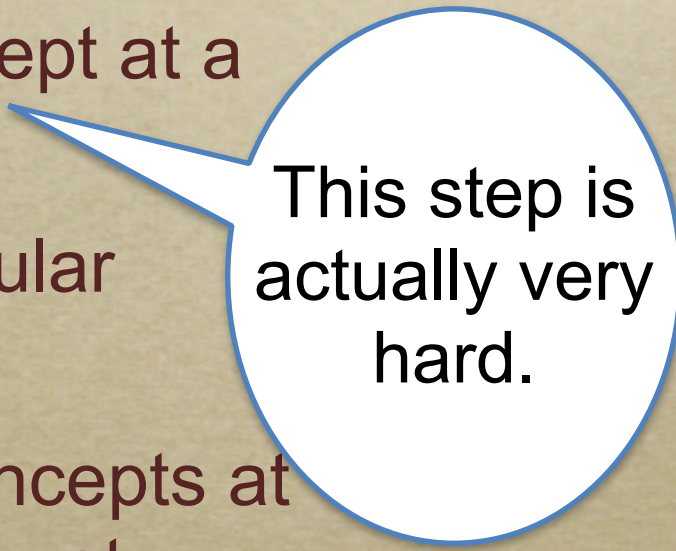
## Calculating Timeliness (Approach 1):

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

# Calculating Concept Difficulty

## Calculating Timeliness (Approach 1):

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

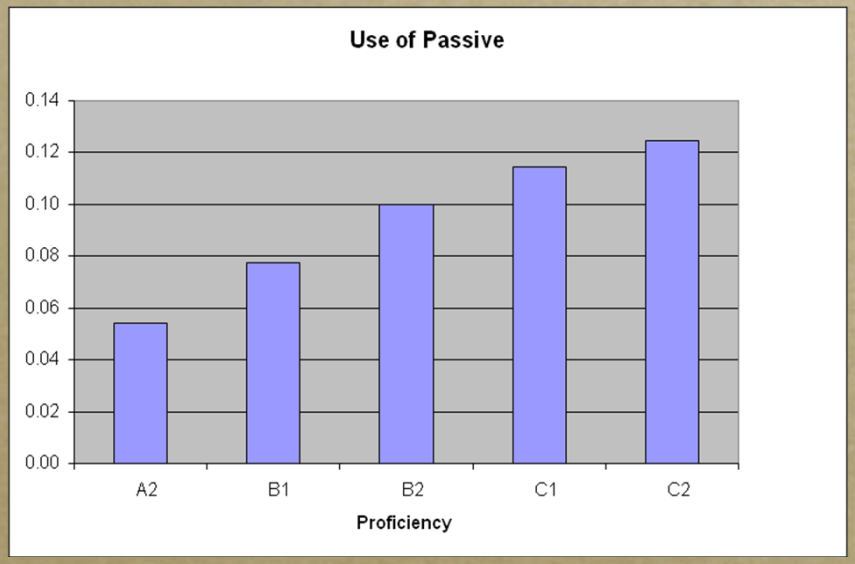


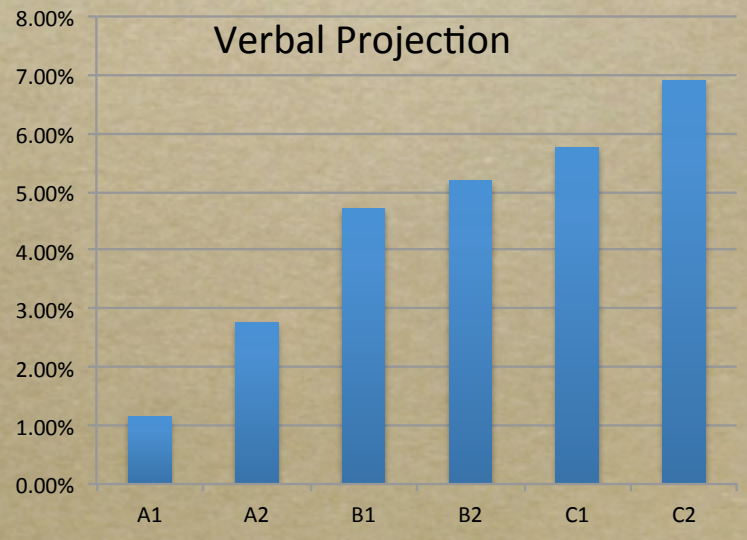
# Calculating Concept Difficulty

## 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?



# Calculating Concept Difficulty

## Calculating Timeliness (Approach 2):

1. Order grammatical concepts in terms of difficulty.
2. Identify the concepts that the learner has assimilated.
3. Timely concepts are the unassimilated concepts of lowest difficulty.

# Calculating Concept Difficulty

## Calculating Timeliness (Approach 2):

1. Order grammatical concepts in terms of difficulty.
2. Identify the concepts that the lea assimilated.
3. Timely concepts are the unassimila concepts of lowest difficulty.

A lot easier than placing features at proficiency levels

# Calculating Concept Difficulty

## Calculating Timeliness (Approach 2):

1. Order grammatical c… difficulty.
2. Identify the concepts that the learner has assimilated.
3. Timely concepts are t… concepts of lowest c…

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

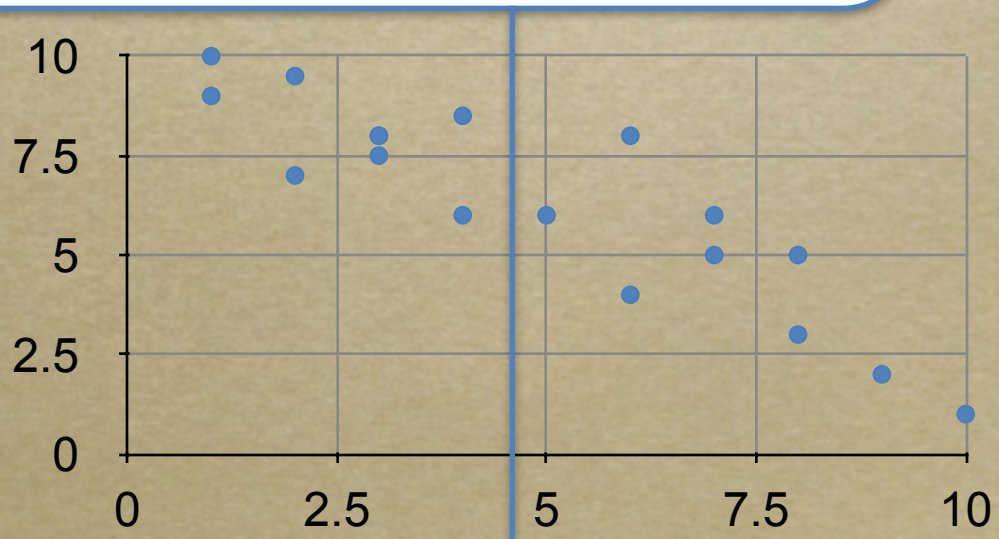
Analysis of their submitted writing to identify successful and unsuccessful applications of the concept

# Calculating Concept Difficulty

## How to order features in difficulty: Using error data:

- For each error type:
  1. Identify all instance of the error
  2. Assign each error 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

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"

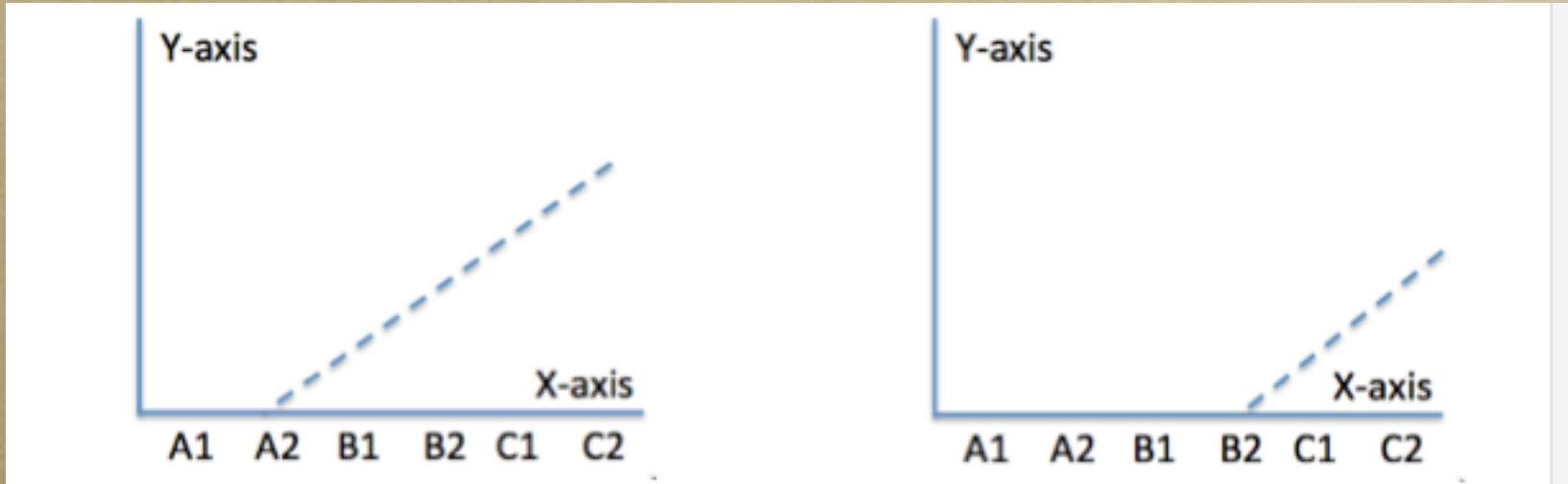
# Calculating Concept Difficulty

## 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 |

# Summary

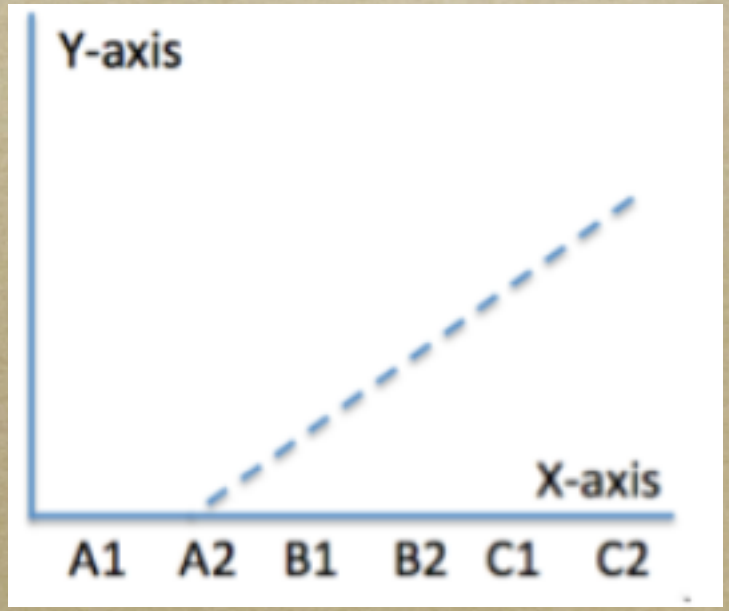
We can derive from our learner corpus the resources we need:

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

# Part 3:
# Contexts of Use

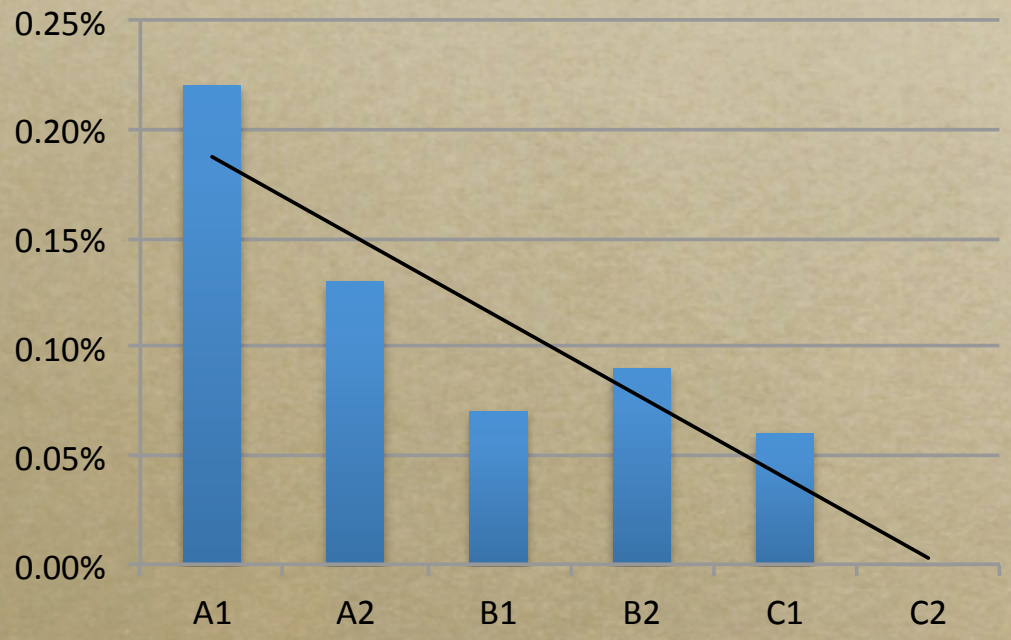
# Contexts of Use

- My initial conception of how learners progressed was that:
  - Learner does't know how to form the structure
  - Learner learns how to form the structure.
  - Learners start to introduce the structure into their production



# Contexts of Use

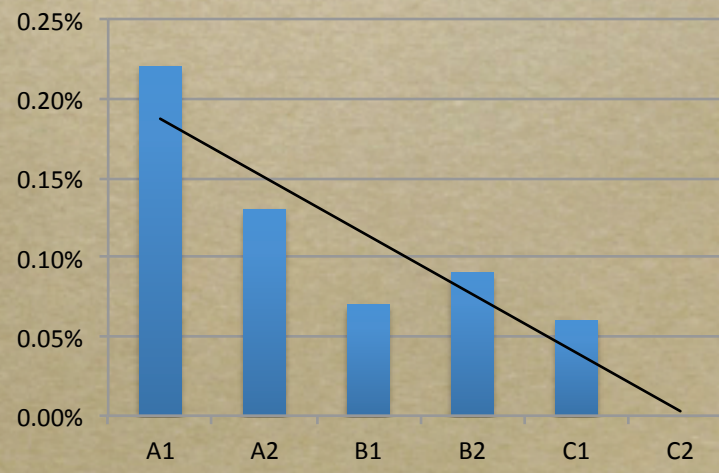
- However, this naive approach fails to explain acquisitional patterns such as:




Use of Past-progressive with rising proficiency

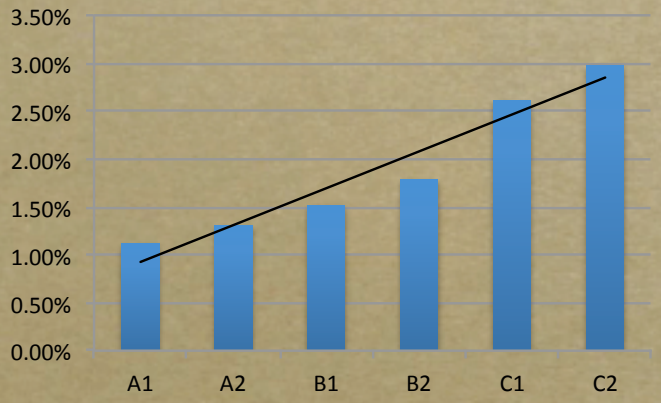
# Contexts of Use

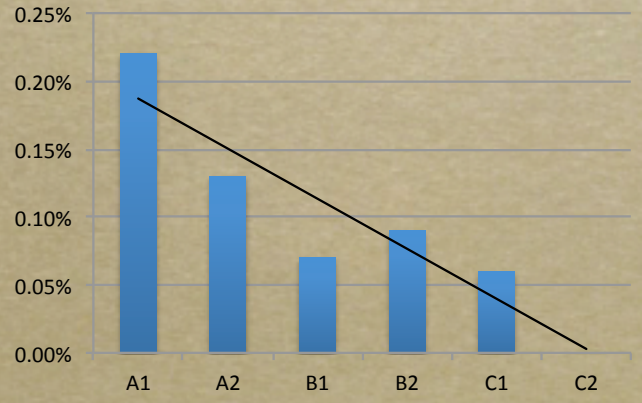
- This common pattern suggests that learners can easily transfer the structure from their mother tongue (in this case, Spanish)
- Learners then learn that the structure is not appropriate for all the contexts they use it in.
- E.g., It seems that (continental) Spanish speakers use the past-progressive in contexts where English natives would use simple-past.



# Contexts of Use

- A revised conception of how learners progressed is:
  - Learner acquires the structure, possibly by transfer from their mother tongue.
  - Learners start to introduce the structure into their production using the same contexts of use as in their mother tongue equivalent.
  - Learner gradually learns in which contexts the structure is (in)appropriate in English.



# Contexts of Use: Present-Perfect

- In English, the primary context of use of present-perfect is to indicate that some past action still has consequence in the present:

  *I have spilt my coffee.*

  *I have eaten already.*

In Continental Spanish (at least in most of the country), present-perfect is a common way to refer to the recent past: what has happened <u>today</u>, regardless of whether the event still affects the present:

At 10am:
English: *I have eaten breakfast*
Spanish: *He desayunado*

At 6pm:
English: *I ate breakfast this morning*
Spanish: *He desayunado esta mañana*

# Contexts of Use: Present-Perfect

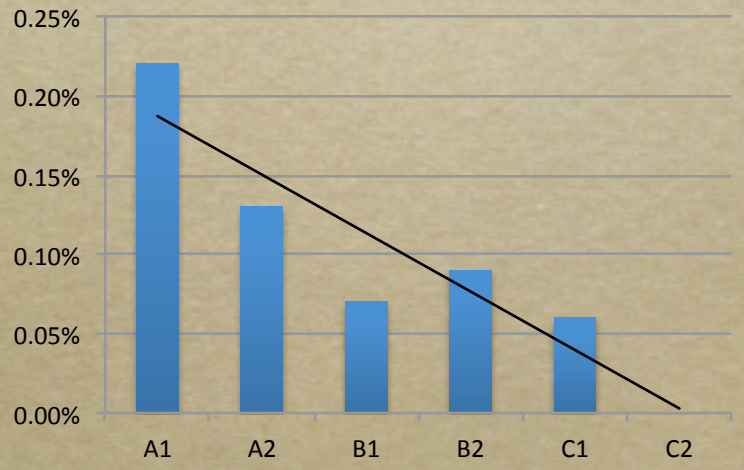
| *Context of Use* | *English* | *Standard Spanish* |
|---|---|---|
| **Past even with current consequence** | *I have broken my arm* | *X* |
| **Past event same day, no consequence implied** | *X*<br>*I ate breakfast this morn…* | *He desayunado esta mañana* |
| **Counting results in a still-open period** | *We have built 20 houses so far this year.* | *Hemos construido 20 casas…* |
| **Life Achievements** | *I have lived in London.* | *He vivido en Londres* |
| **Specifying first time** | *This is the first time I've eaten Sushi.* | *X*<br>*… que como arroz* |
| **Specifying length of continuing action** | *I have played tennis for 10 years* | *X* |

# Contexts of Use: Past-Progressive

- Spanish speakers often use the past-progressives in their L1 in contexts where we don't in English.
- In English, we use it in contexts where we will then relate what happened within that continuous action.
- Not necessary so in Spanish

| | |
|---|---|
| ***Spanish:*** | *Estaba hablando con Susana ayer.* |
| ***Lit:*** | ***(I) was talking with Susana yesterday*** |
| ***English Equiv.*** | ***I talked with Susana yesterday*** |

- As a result, Spanish learners of English over-produce past-progressives until they master the contexts of use.




*Use of past-progressive aspect*

# Contexts of Use: Articles

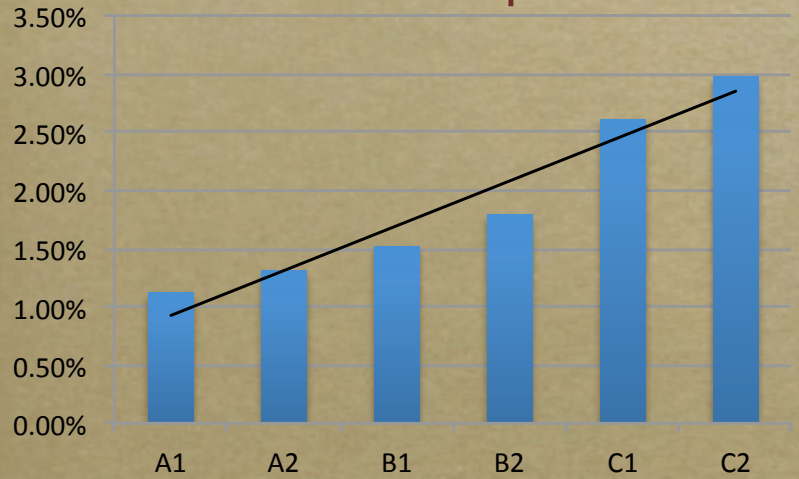
- The most frequent source of 'syntactic' error in our learners relates to:
  - Producing an article when one is not appropriate:
    *~~The~~ drugs are a problem for ~~the~~ society.*
  - Not producing an article when one is appropriate
    *…in the first semester of () year.*

While previously we saw these as errors of form, now we see there as errors of context of use: particular contexts of use require an article, others do not.

| | English | Spanish | Example |
|---|---|---|---|
| Specific: recoverable | the | el/la | *the water* |
| Specific: non-recov. single | a/an | un/una | *a dog* |
| Specific: non-recoverable plural | some/Ø | unos/unas | *some dogs/dogs* |
| Specific: non-recoverable non-countable | some/Ø | Ø | *some water/water*<br>*some doubt/doubt* |
| Generic: singular | a/an | un/una | *a cat* |
| Generic: plural (i) | Ø | los/las | *cats/los gatos* |
| (ii) | some | unos/unas | *some cats/unos gatos* |
| Generic: non-countable | Ø | el/la | *society/la sociedad* |
| Exception "workplace"<br>(a) 'home' | Ø | Ø | *I went home.*<br>*fui a casa* |
| Exception "workplace"<br>(b) work, prison, school | Ø | el/la | *I went to work/school*<br>*fui al trabajo (al= a el)*<br>*fui a la escuela* |
| Exception "meals" | Ø | el/la | *breakfast/el desayuno* |
| Exception: "percent" | Ø | el/la | *20% of… / el 20% de* |

# Contexts of Use: Passive

- Spanish has two passive forms:
  - 'ser' (='be') passive (equivalent construction to the English passive:
    - *Juan **está** construyendo una casa*
    - *Juan is building a house*
  - 'se'-passive: *Se venden huevos. (Eggs are sold here)*
- In Spanish, many verbs do not work well with the 'ser' passive.
- As a result, when starting to speak/write English, learners use active voice where a native would have used a passive.
- As learners progress, they learn which English verbs allow passive, and thus start to produce them more.



# Contexts of Use: Consequences for our study

**Entire study of syntactic errors needs to be revised:**

- Many errors classed as syntactic errors are not truly syntactic, rather productions of syntactically valid structures in the wrong context of use.
- E.g., article errors:

  ***The*** *drugs are a problem for* ***the*** *society*

  - Coded as*: article-present-not-required*
  - *New coding:*
    - inappropriate use of definite article for generic reference
      (a contextual error, not a syntactic error)

# Contexts of Use

**Consequence**: We are in the process of recoding our met critical syntactic errors to reflect:

- the difference between errors of form and errors of context of use.
- In the case of errors of use, we code also the particular contextual feature which is broken:
- E.g, *They have <u>much</u> money*.
  - much-used-in-positive-declarative-context
- E.g, *The drugs are a problem for …*
  - genetic-plurals-dont-take-definite-article
- E.g, *I was lunching with Mary yesterday*
  - past-prog-used-where-simple-past-adequate

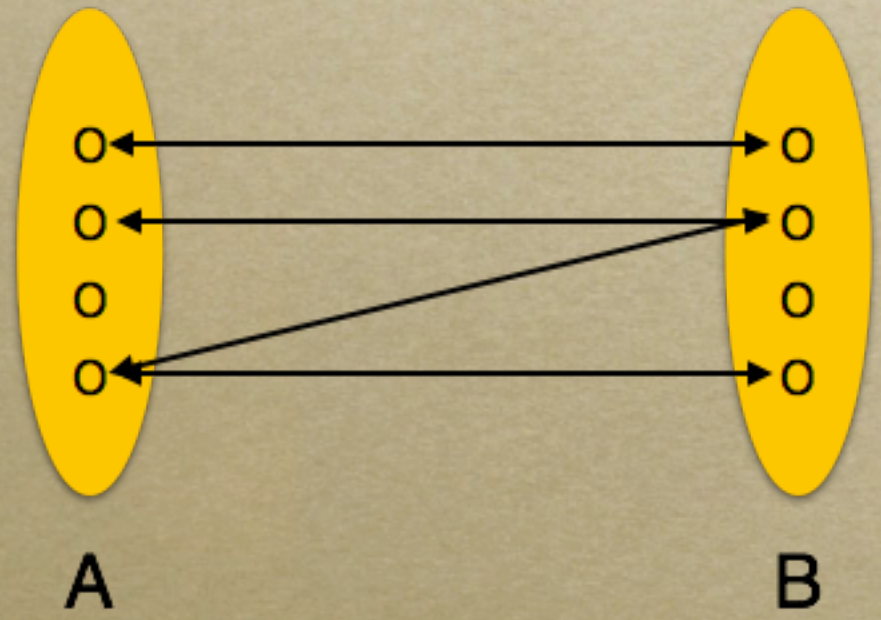
# Contexts of use: Problems for our approach

- **Syntactic Analysis**: Automatic recognition of incorrect contexts of use not always possible:
  - Easy: *I have much money*
  - Difficult:
    - *The drugs are a problem for society.*
    - *I have eaten breakfast this morning* (when context makes clear this is later in the day)

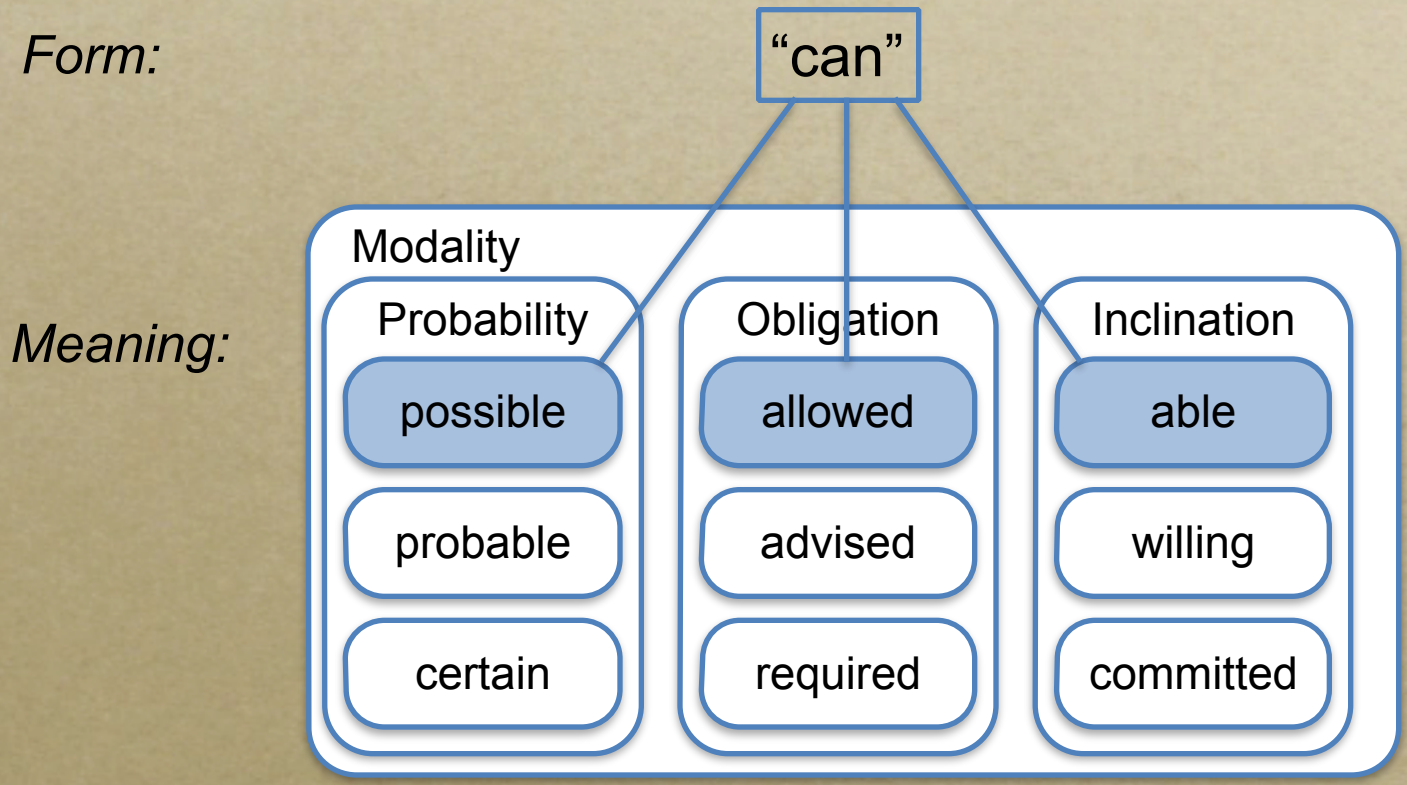
# Contexts of use: Partial solution

- Automatic identification of changing degree of use over rising proficiency reveal potential cases that need to be addressed.
- Manual exploration of these areas to reveal which contexts of use are giving problems to our learners.
- Materials put into our Question database to help identify learner-performance in these contexts.

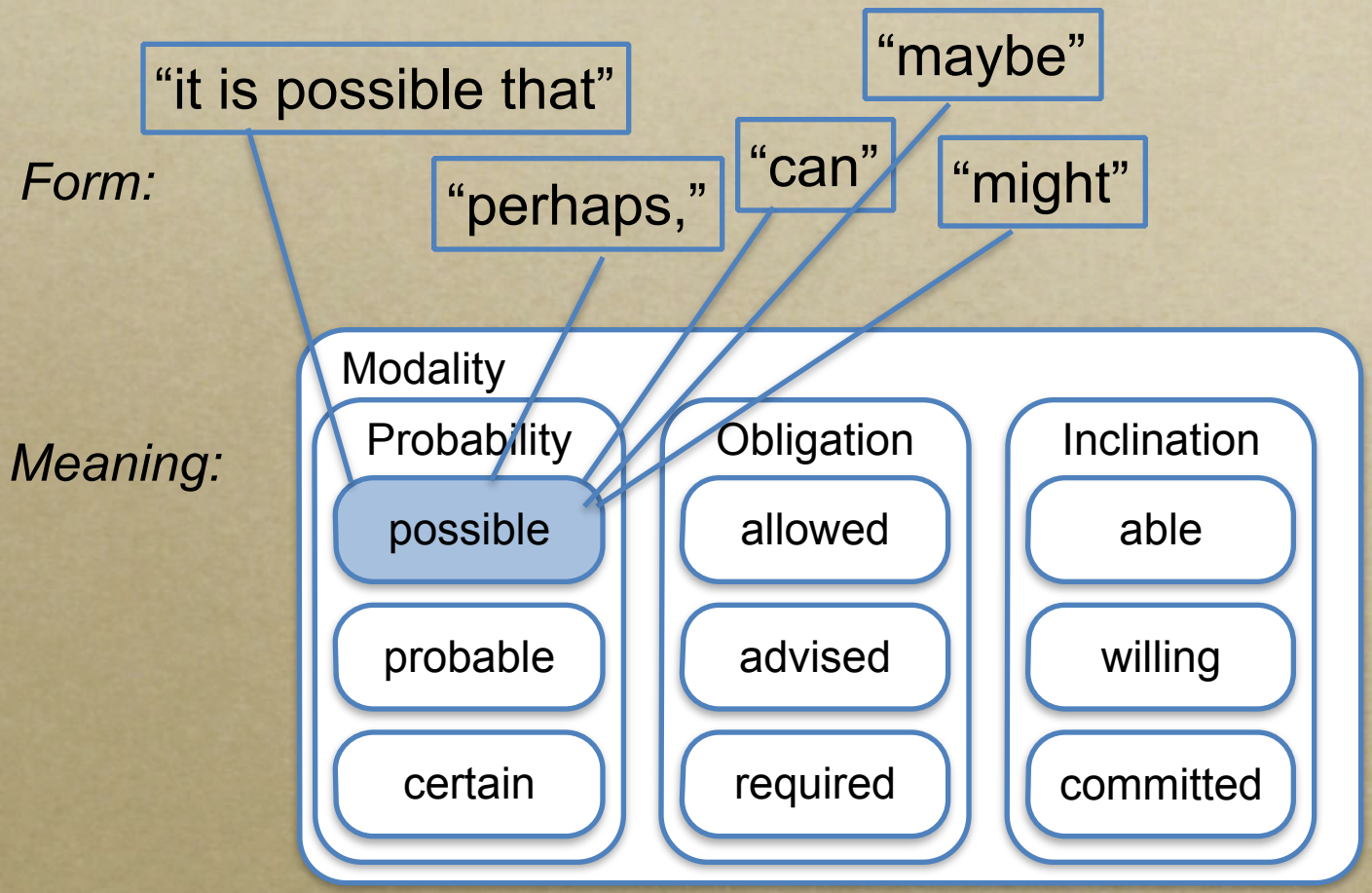
# Part 4:
# The many-to-many mappings between contexts and forms



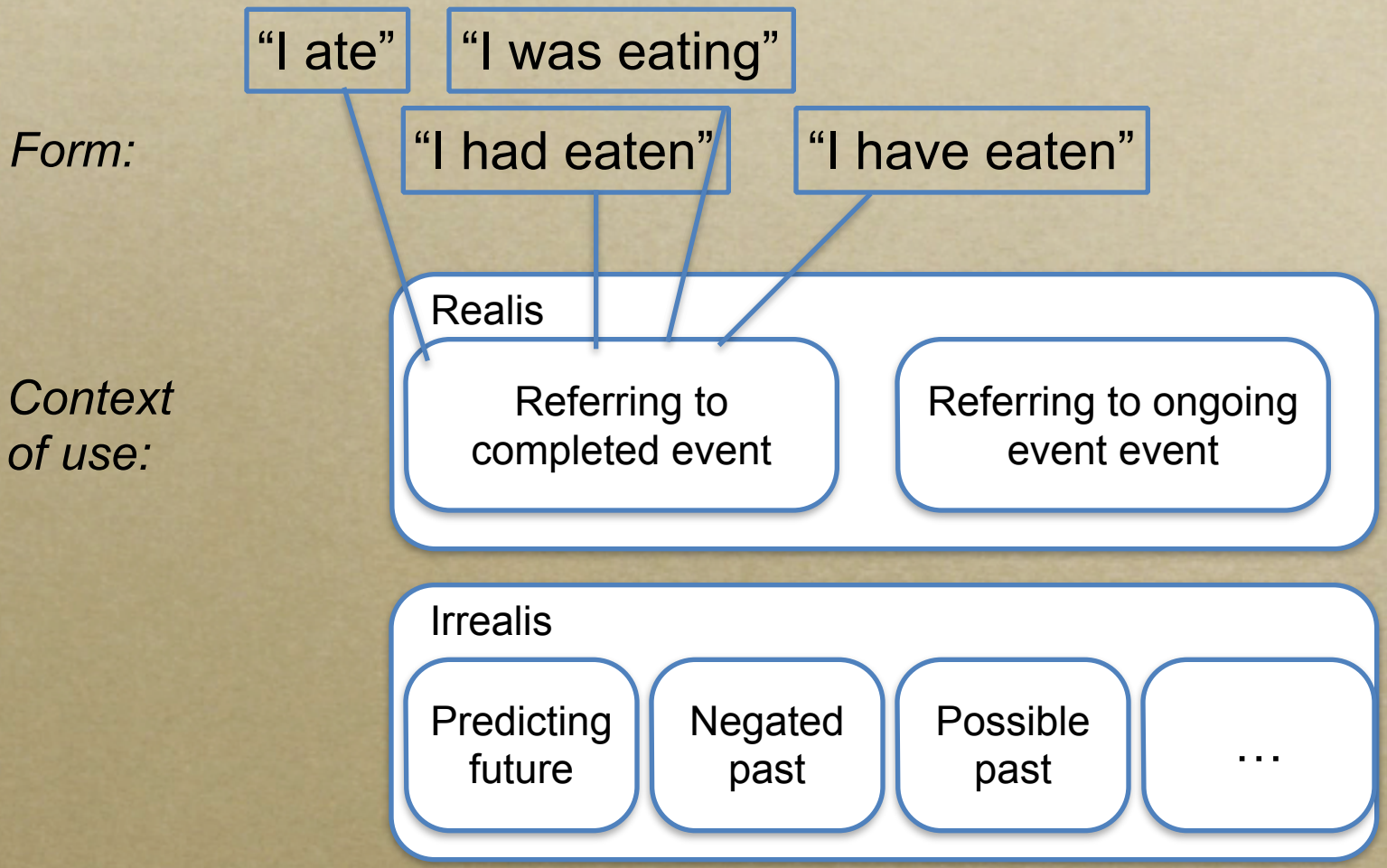
# Forms have multiple contexts of use
(in this case, context of use = 'to express a particular meaning')



A given meaning can be grammaticalised in many different ways:



# With tense aspect:



# Possibilities for learner writer feedback system

System
- Identifies form
- Guess intended context of use
- Offer alternative forms appropriate for this context

I went home and was eating breakfast.
Then I did some study.

Alternatives:
- "ate breakfast"
- "had eaten breakfast"
- "have had breakfast"

# Summary and Conclusions

# Summary

- While some structures can be transferred fairly directly from the mother tongue, we need to be aware that learning a structure involves both:
  - knowing HOW to produce the structure,
  - knowing WHEN it can be appropriately used.

- Online learning systems need to be designed to deal with both kinds of knowledge.