Error Analysis in the TREACLE project

Mick O'Donnell, Susana Murcia, Rebeca Garcia

Universidad Autónoma de Madrid

Contents

1. The TREACLE Project

- A cooperation between **Universidad Autonoma de Madrid** and **University Politecnica de Valencia** (Penny McDonald, Keith Stuart, Maria Boquera)
- Funded by Ministerio de Ciencia e Innovación 2010-2012 (FFI2009-14436/FILO)
- Goals:
  - To map at what **proficiency level** each grammatical structure is best taught in a Spanish context.
  - Adjust the grammar teaching **syllabus** at our universities in line with these results.
  - Use the error corpus as a **resource** for **teaching examples** and **online exercises**.
  - **Automatic proficiency assessment** based on structures and errors in student texts.

1.1 The Corpus

- The project involves two corpora:

  The **WriCLE** corpus (UAM) - Written Corpus of Learner English. 700,000 words, written by Spanish learners of English at University level. Rollinson and Mendikoetxea (2009).

  The **UPV Learner Corpus** (UPV) containing 150,000 words of shorter texts by ESP students.

- ONLY A 18,000 word subcorpus error annotated so far (28 texts)

2. Role of Error Analysis in the Project

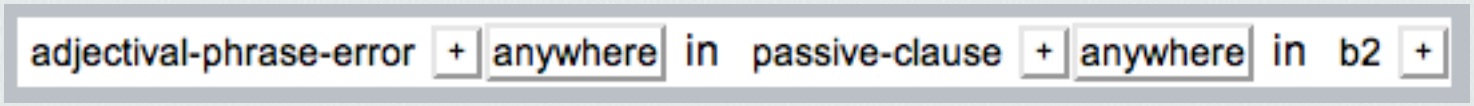
- Error analysis is one way to explore the grammatical competence of students at each level (e.g. Dagneaux et al 1998).
- However, some students make few errors, because they avoid structures they are not sure about
- More adventurous students take risks and thus make more errors.
- We thus take a two-pronged approach:
  - Automatic syntactic tagging of corpus to see what structures students are attempting;
  - Manual error analysis to see what they do wrong.
- Only both together give the full picture.

3. Software for Error Analysis

- We use (and develop) UAM CorpusTool, software for text annotation
- Multi-layer annotation of a corpus (e.g., we use 3 layers: Document, Grammar, and Error)
- User provides annotation schemes (tags organised into a tag heirarchy) using graphical editor.
- Some schemes built in and optionally available (Error, English Syntax, Appraisal Analysis)

3. Software for Error Analysis

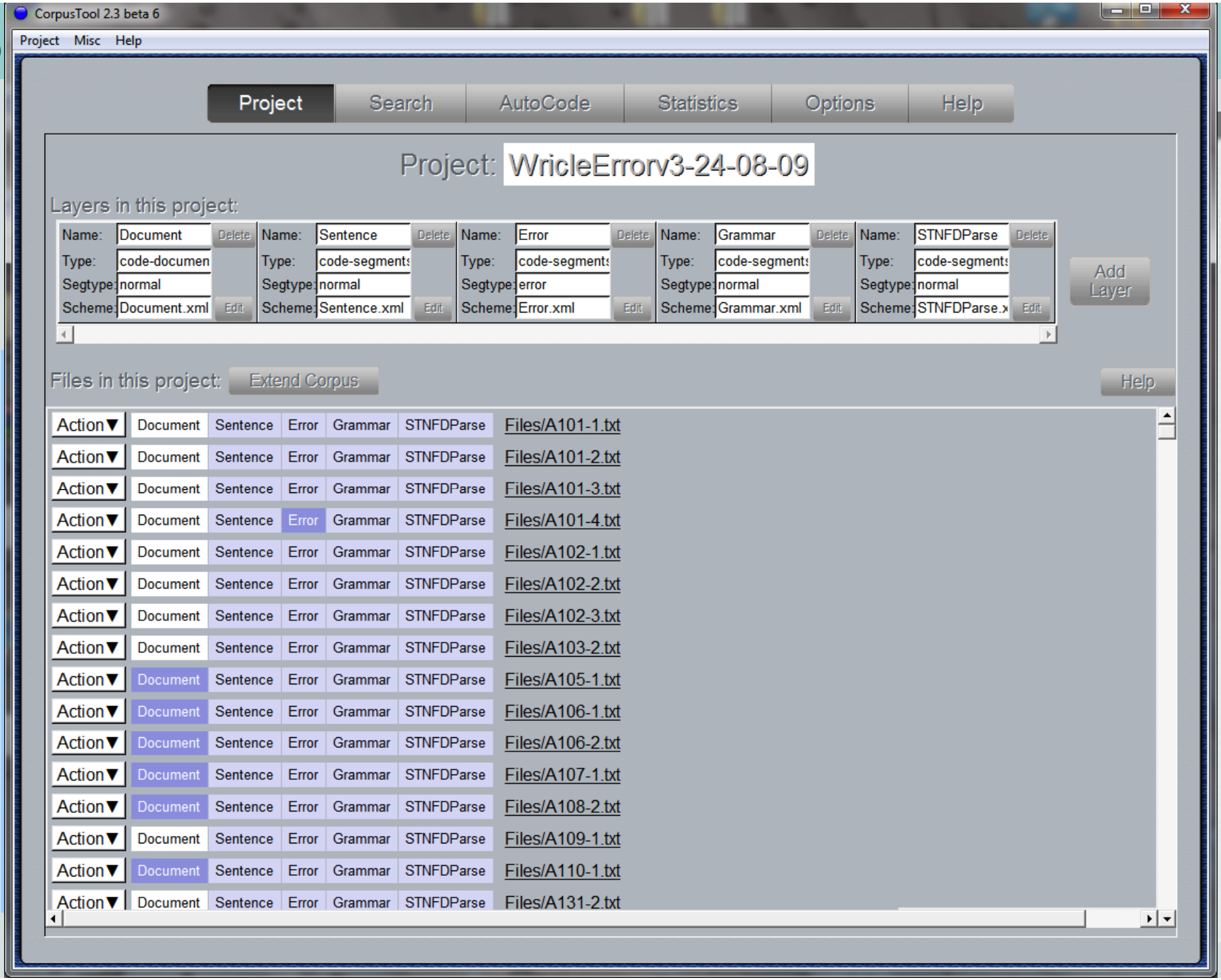
- Tool provides cross-level search facilities

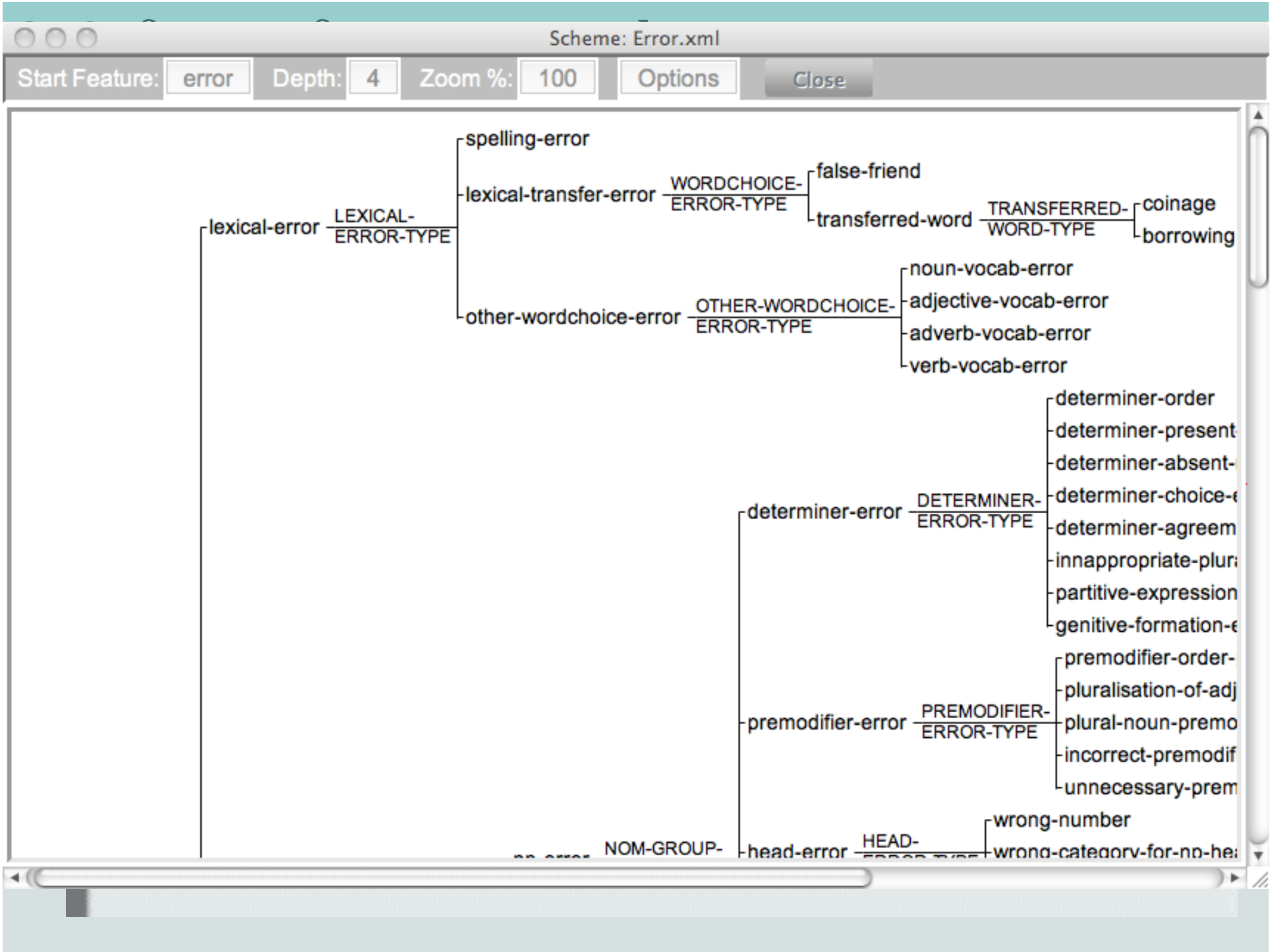


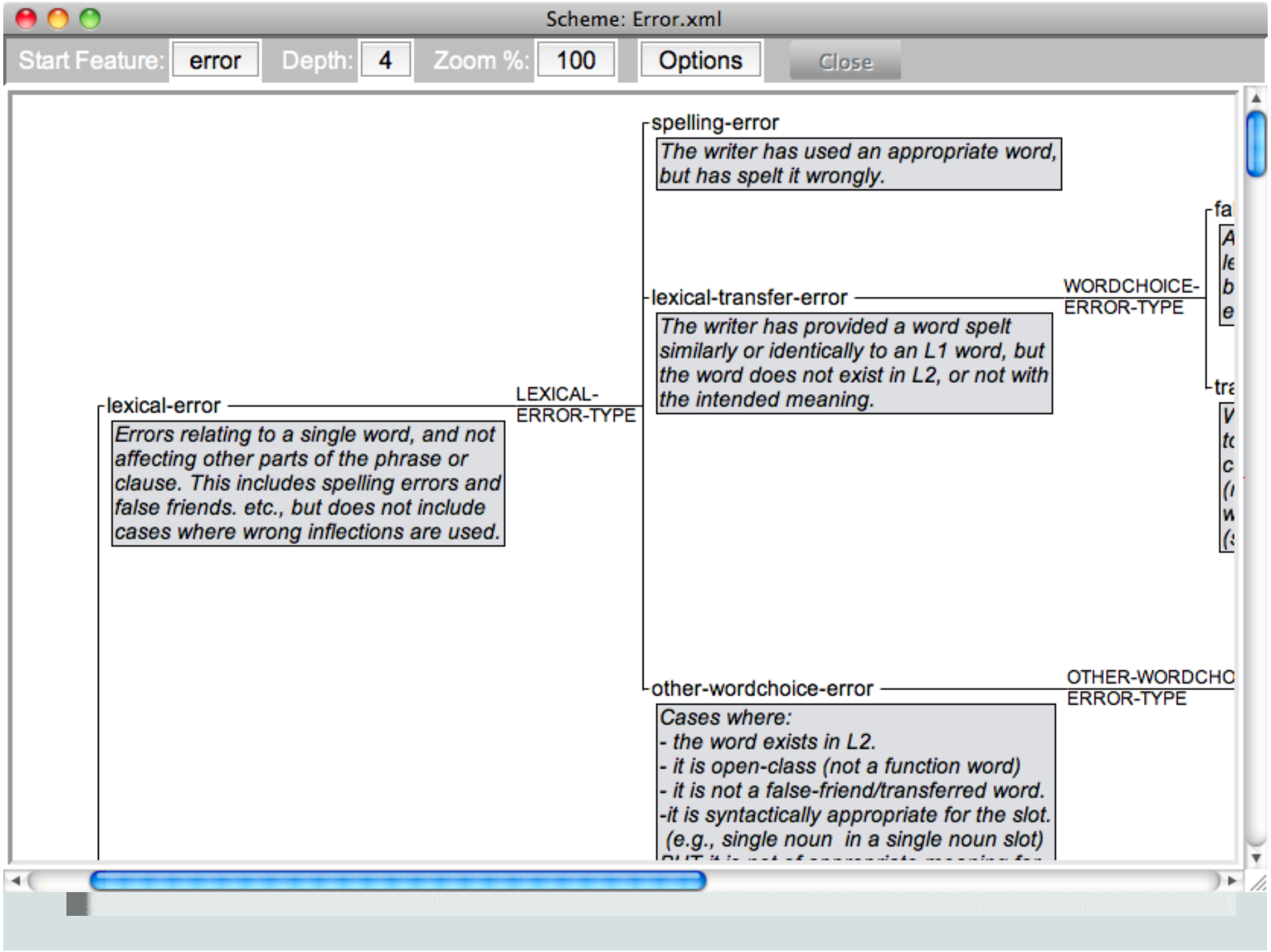
- Tool provides statistical reports (compare two subsets etc.)

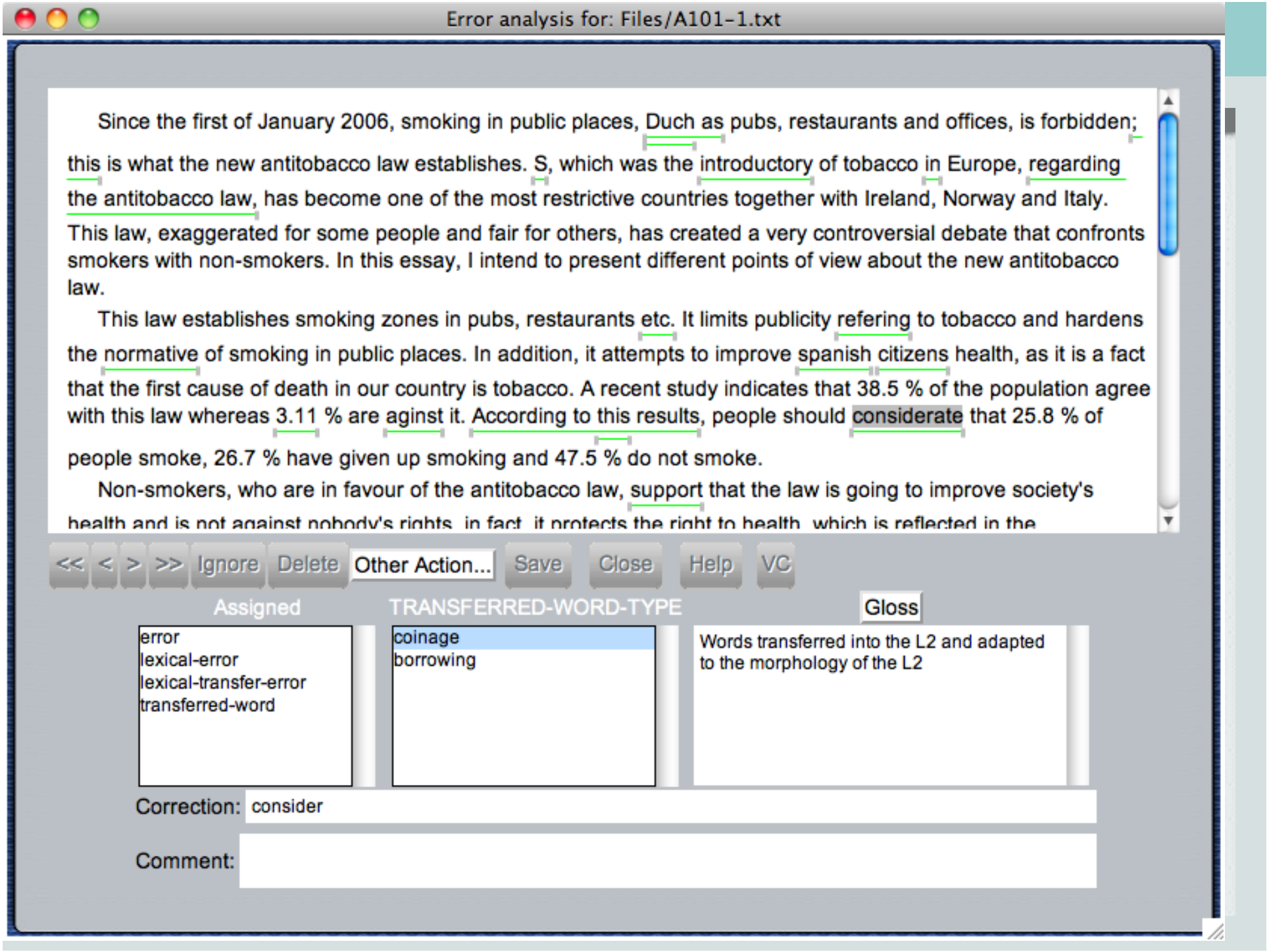
- Available for free (Mac, PC) from:

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

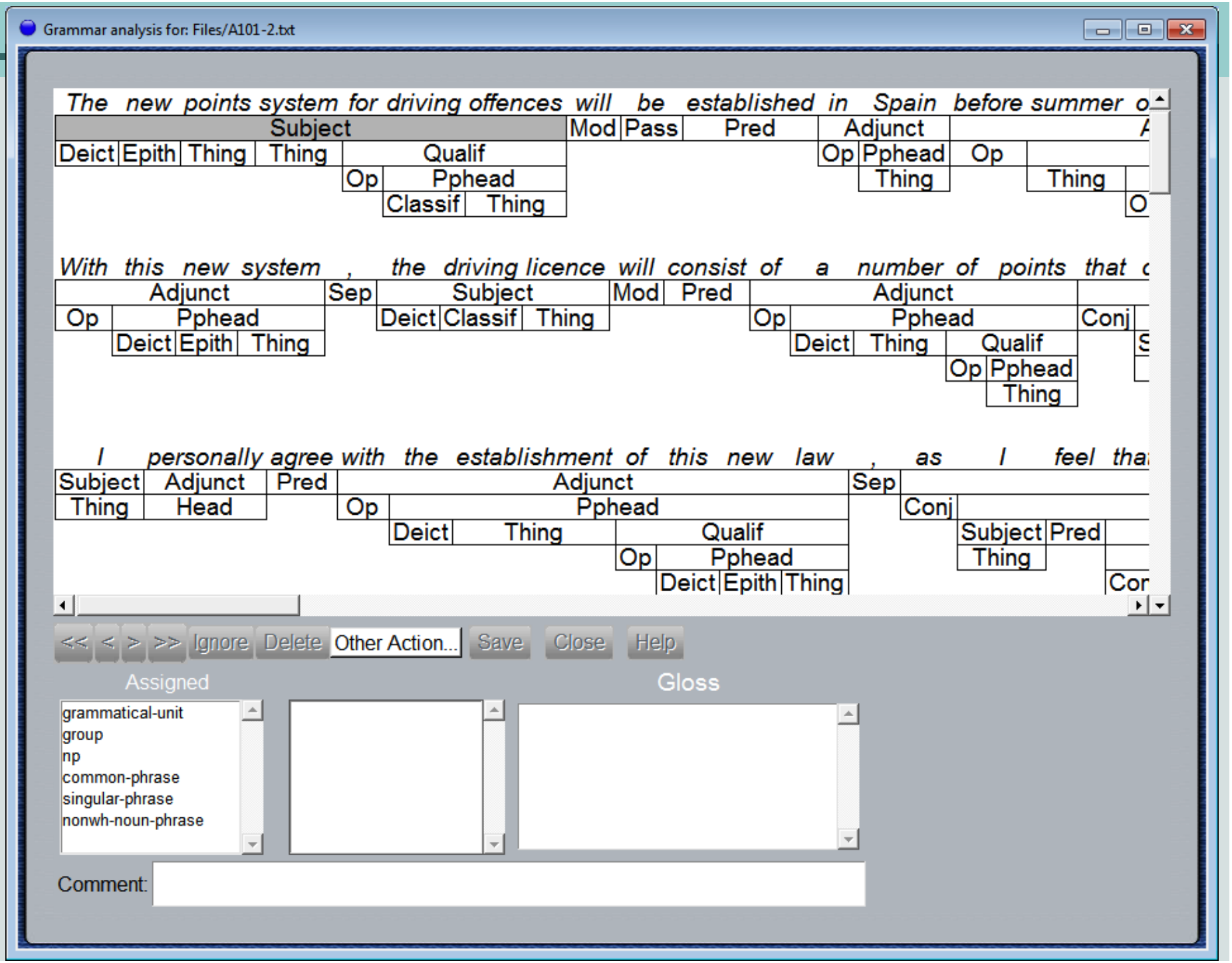

CorpusTool 2.3 beta 6
Project Misc Help
Project Search AutoCode Statistics Options Help
Project: WricleErrorv3-24-08-09
Layers in this project:
Name: Document Delete
Type: code-documen
Segtype: normal
Scheme: Document.xml Edit
Name: Sentence Delete
Type: code-segments
Segtype: normal
Scheme: Sentence.xml Edit
Name: Error Delete
Type: code-segments
Segtype: error
Scheme: Error.xml Edit
Name: Grammar Delete
Type: code-segments
Segtype: normal
Scheme: Grammar.xml Edit
Name: STNFDParse Delete
Type: code-segments
Segtype: normal
Scheme: STNFDParse.x Edit
Add Layer
Files in this project:
Extend Corpus
Help
Action▼ Document Sentence Error Grammar STNFDParse Files/A101-1.txt
Action▼ Document Sentence Error Grammar STNFDParse Files/A101-2.txt
Action▼ Document Sentence Error Grammar STNFDParse Files/A101-3.txt
Action▼ Document Sentence Error Grammar STNFDParse Files/A101-4.txt
Action▼ Document Sentence Error Grammar STNFDParse Files/A102-1.txt
Action▼ Document Sentence Error Grammar STNFDParse Files/A102-2.txt
Action▼ Document Sentence Error Grammar STNFDParse Files/A102-3.txt
Action▼ Document Sentence Error Grammar STNFDParse Files/A103-2.txt
Action▼ Document Sentence Error Grammar STNFDParse Files/A105-1.txt
Action▼ Document Sentence Error Grammar STNFDParse Files/A106-1.txt
Action▼ Document Sentence Error Grammar STNFDParse Files/A106-2.txt
Action▼ Document Sentence Error Grammar STNFDParse Files/A107-1.txt
Action▼ Document Sentence Error Grammar STNFDParse Files/A108-2.txt
Action▼ Document Sentence Error Grammar STNFDParse Files/A109-1.txt
Action▼ Document Sentence Error Grammar STNFDParse Files/A110-1.txt
Action▼ Document Sentence Error Grammar STNFDParse Files/A131-2.txt


Scheme: Error.xml
Start Feature: error
Depth: 4
Zoom %: 100
Options
Close
error
lexical-error
LEXICAL-ERROR-TYPE
spelling-error
lexical-transfer-error
WORDCHOICE-ERROR-TYPE
false-friend
transferred-word
TRANSFERRED-WORD-TYPE
coinage
borrowing
other-wordchoice-error
OTHER-WORDCHOICE-ERROR-TYPE
noun-vocab-error
adjective-vocab-error
adverb-vocab-error
verb-vocab-error
determiner-error
DETERMINER-ERROR-TYPE
determiner-order
determiner-present
determiner-absent-
determiner-choice-
determiner-agreem
innappropriate-plur
partitive-expression
genitive-formation-
premodifier-error
PREMODIFIER-ERROR-TYPE
premodifier-order-
pluralisation-of-adj
plural-noun-premo
incorrect-premodif
unnecessary-prem
NOM-GROUP-
head-error
HEAD-
wrong-number
wrong-category-for-np-he


Scheme: Error.xml
Start Feature: error
Depth: 4
Zoom %: 100
Options
Close
lexical-error
Errors relating to a single word, and not affecting other parts of the phrase or clause. This includes spelling errors and false friends. etc., but does not include cases where wrong inflections are used.
LEXICAL-ERROR-TYPE
spelling-error
The writer has used an appropriate word, but has spelt it wrongly.
lexical-transfer-error
The writer has provided a word spelt similarly or identically to an L1 word, but the word does not exist in L2, or not with the intended meaning.
WORDCHOICE-ERROR-TYPE
other-wordchoice-error
Cases where:
- the word exists in L2.
- it is open-class (not a function word)
- it is not a false-friend/transferred word.
-it is syntactically appropriate for the slot.
(e.g., single noun in a single noun slot)
OTHER-WORDCHO
ERROR-TYPE


Error analysis for: Files/A101-1.txt
Since the first of January 2006, smoking in public places, Duch as pubs, restaurants and offices, is forbidden; this is what the new antitobacco law establishes. S, which was the introductory of tobacco in Europe, regarding the antitobacco law, has become one of the most restrictive countries together with Ireland, Norway and Italy. This law, exaggerated for some people and fair for others, has created a very controversial debate that confronts smokers with non-smokers. In this essay, I intend to present different points of view about the new antitobacco law.
This law establishes smoking zones in pubs, restaurants etc. It limits publicity refering to tobacco and hardens the normative of smoking in public places. In addition, it attempts to improve spanish citizens health, as it is a fact that the first cause of death in our country is tobacco. A recent study indicates that 38.5 % of the population agree with this law whereas 3.11 % are aginst it. According to this results, people should considerate that 25.8 % of people smoke, 26.7 % have given up smoking and 47.5 % do not smoke.
Non-smokers, who are in favour of the antitobacco law, support that the law is going to improve society's
<< < > >> Ignore Delete Other Action... Save Close Help VC
Assigned
error
lexical-error
lexical-transfer-error
transferred-word
TRANSFERRED-WORD-TYPE
coinage
borrowing
Gloss
Words transferred into the L2 and adapted to the morphology of the L2
Correction: consider
Comment:

3.


Grammar analysis for: Files/A101-2.txt
The new points system for driving offences will be established in Spain before summer o
Subject
Mod
Pass
Pred
Adjunct
Deict
Epith
Thing
Thing
Qualif
Op
Pphead
Classif
Thing
Op
Pphead
Thing
Op
Thing
With this new system , the driving licence will consist of a number of points that c
Adjunct
Sep
Subject
Mod
Pred
Adjunct
Op
Pphead
Deict
Epith
Thing
Deict
Classif
Thing
Op
Pphead
Conj
Deict
Thing
Qualif
Op
Pphead
Thing
I personally agree with the establishment of this new law , as I feel tha
Subject
Adjunct
Pred
Adjunct
Sep
Thing
Head
Op
Pphead
Conj
Deict
Thing
Qualif
Subject
Pred
Thing
Op
Pphead
Deict
Epith
Thing
Cor
<< < > >> Ignore Delete Other Action... Save Close Help
Assigned
grammatical-unit
group
np
common-phrase
singular-phrase
nonwh-noun-phrase
Gloss
Comment:

3.5 Data Representation

- XML - machine readable by other projects
- STANDOFF ANNOTATION: allows multiple annotation layers for each text file.

```
<document>
 <header><textfile>Files/A101-3.txt</textfile></header>
 <segments>
   <segment id="44" start="11" end="16"
     features="error;lexical-error;spelling-error"
     state="active" correction="Mayor" />
   <segment id="45" start="77" end="86"
     features="error;lexical-error;spelling-error"
     state="active" correction="vehicles" />
 ....
```

- Beginning and end of all errors is recorded (Leuven aproach just records start in general)

4. Basic Principles of Error Coding in Treacle

Basic PHILOSOPHY

- The primary criteria behind the error scheme is to allow errors to be related to the English grammar teaching syllabus (Quirk and Greenbaum model assumed)
- We are thus not interested in lexically organized "dictionaries of errors"
  - Rather, we focus on the grammatical topic in which the error would be taught.
- We also avoid connecting errors to word classes (e.g. adverb error)
  - Rather, we associate them to the grammatical unit which provides the context for the error (phrase or clause)
  - E.g., "He runs quick" is not an adverb error, but rather an error at clause level (innapropriate Adjunct filler).

4. Basic Principles of Error Coding in Treacle

- **Code the text vs. code the correction**: In coding errors, we can code in respect to:
  - what the learner actually writes, or
  - what the corrected text should be.
- For instance, if a learner writes

  a woman beautiful

  …is this a noun premodifier problem? (what should have been a premodifier was placed after the noun)

  Or a postmodifier problem? (incorrect type for a postmodifier)
- In general, we follow the principle: *if there is a conflict, we code in relation to what the learner has written, not to what they should have written*.
- Rationale: there are various possible corrections to some errors, and if we code to the corrected text, the coder's choice of correction determines the error category.

4. Basic Principles of Error Coding in Treacle

- When segmenting errors, we use **minimal segmentation** – only select as much as you need to make the correction, with the exception that you should never select parts of words.
- We don't need to identify whole syntactic units, because the automatic syntactic analysis identifies clause and phrase boundaries.
- Examples:
  - in the other hand
  - These person
  - They advocate immigration fully

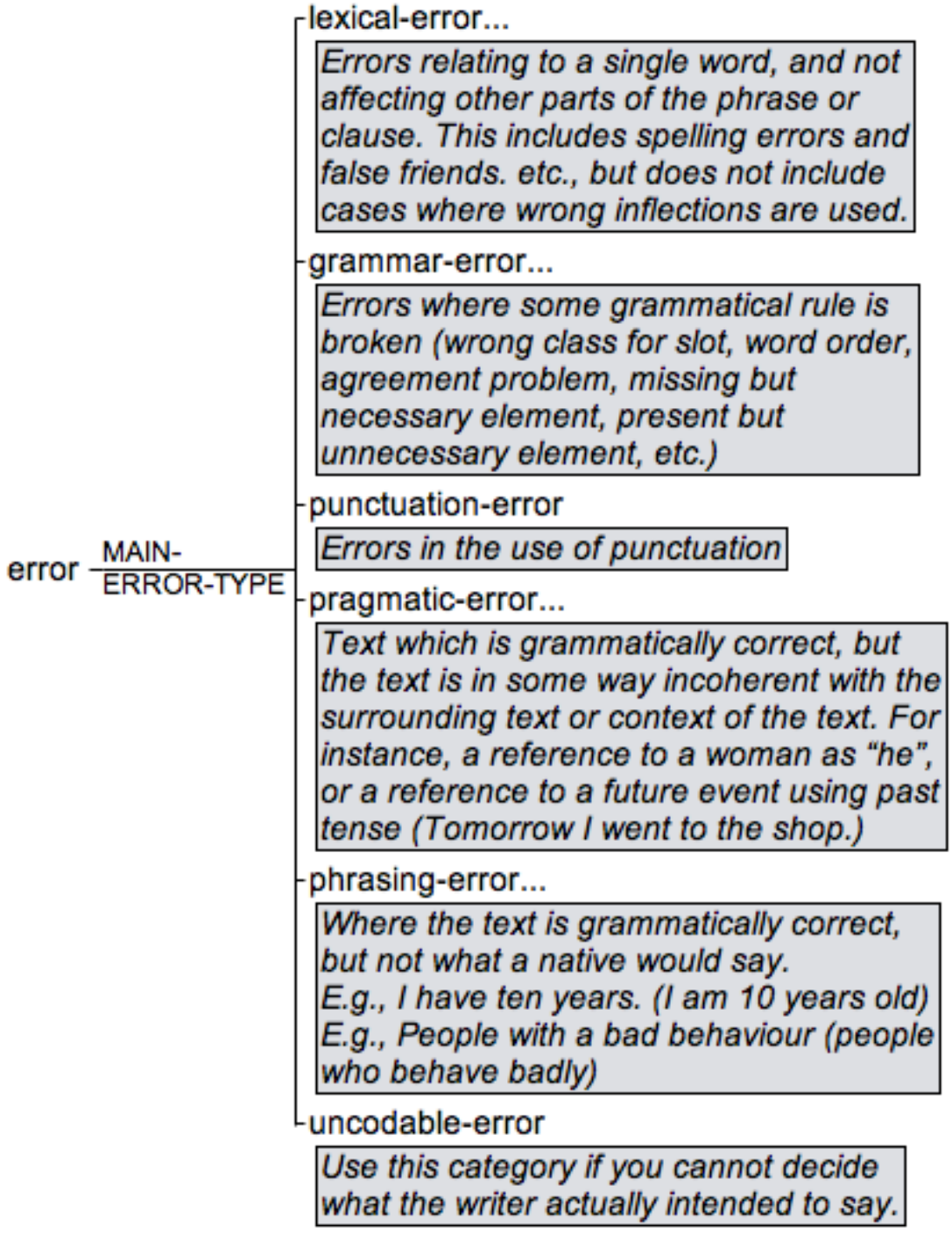
4. Basic Principles of Error Coding in Treacle

The Coding Criteria Document

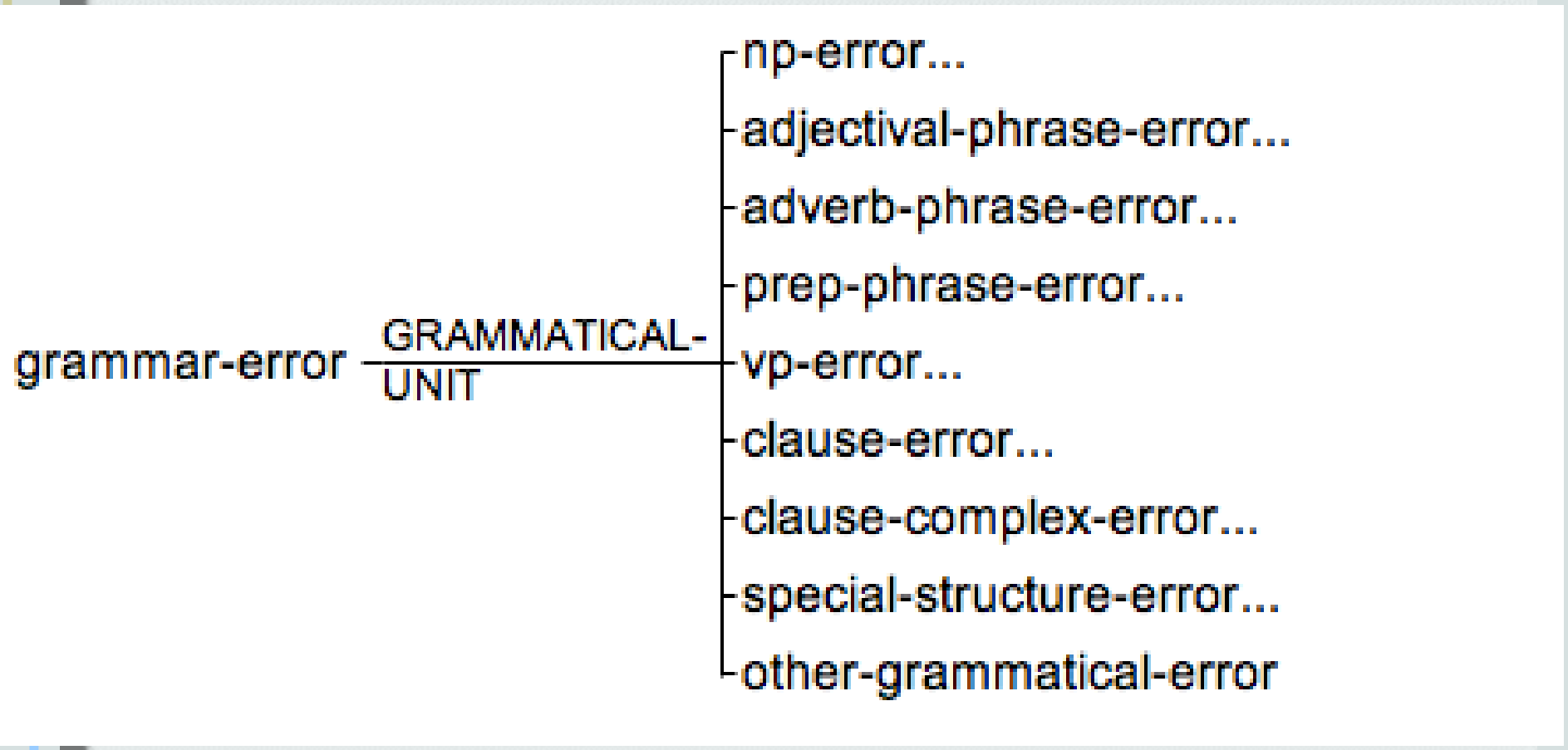
- We maintain a coding criteria document recording all decisions we reach in coding, organised to follow the structure of the coding scheme (20 pages at present)

- Coding criteria are also recorded in the coding scheme so that criteria are visible as one tags errors.

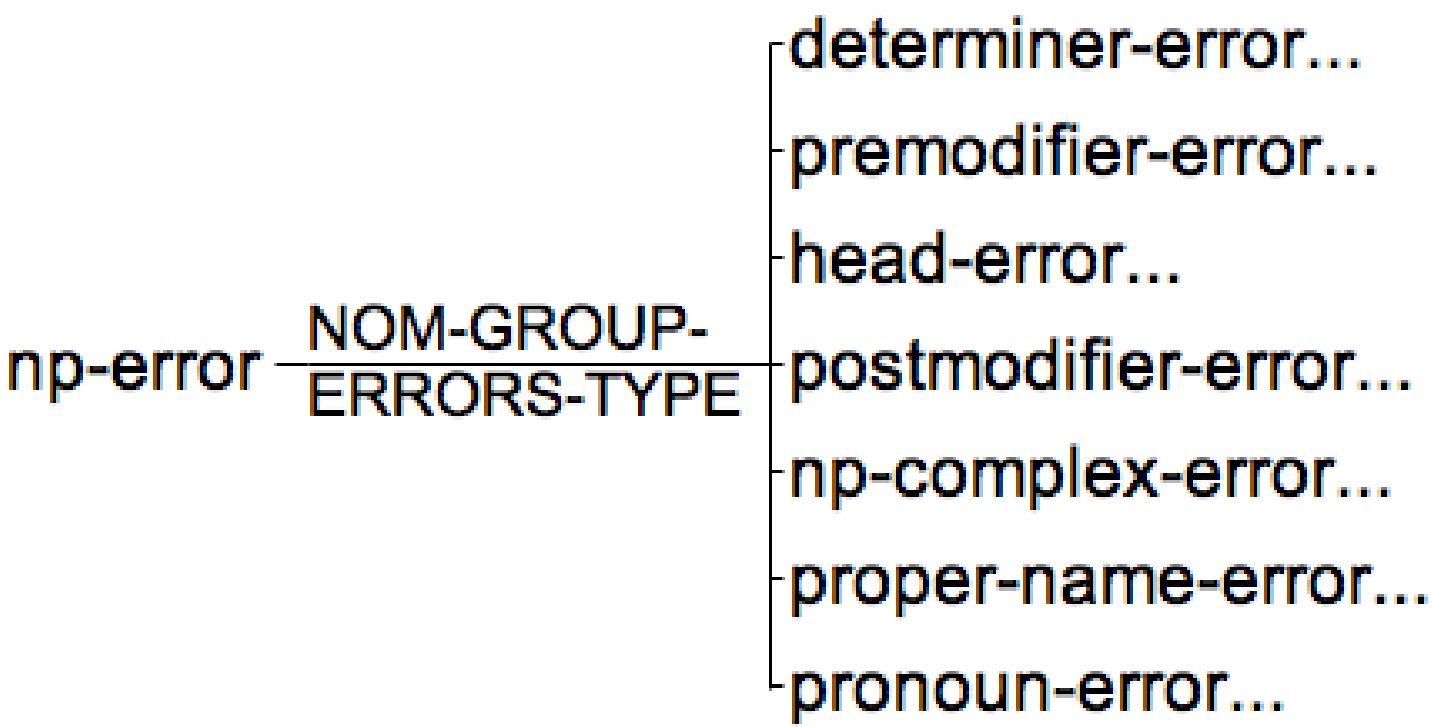
Root


error
MAIN-ERROR-TYPE
lexical-error...
Errors relating to a single word, and not affecting other parts of the phrase or clause. This includes spelling errors and false friends. etc., but does not include cases where wrong inflections are used.
grammar-error...
Errors where some grammatical rule is broken (wrong class for slot, word order, agreement problem, missing but necessary element, present but unnecessary element, etc.)
punctuation-error
Errors in the use of punctuation
pragmatic-error...
Text which is grammatically correct, but the text is in some way incoherent with the surrounding text or context of the text. For instance, a reference to a woman as "he", or a reference to a future event using past tense (Tomorrow I went to the shop.)
phrasing-error...
Where the text is grammatically correct, but not what a native would say.
E.g., I have ten years. (I am 10 years old)
E.g., People with a bad behaviour (people who behave badly)
uncodable-error
Use this category if you cannot decide what the writer actually intended to say.

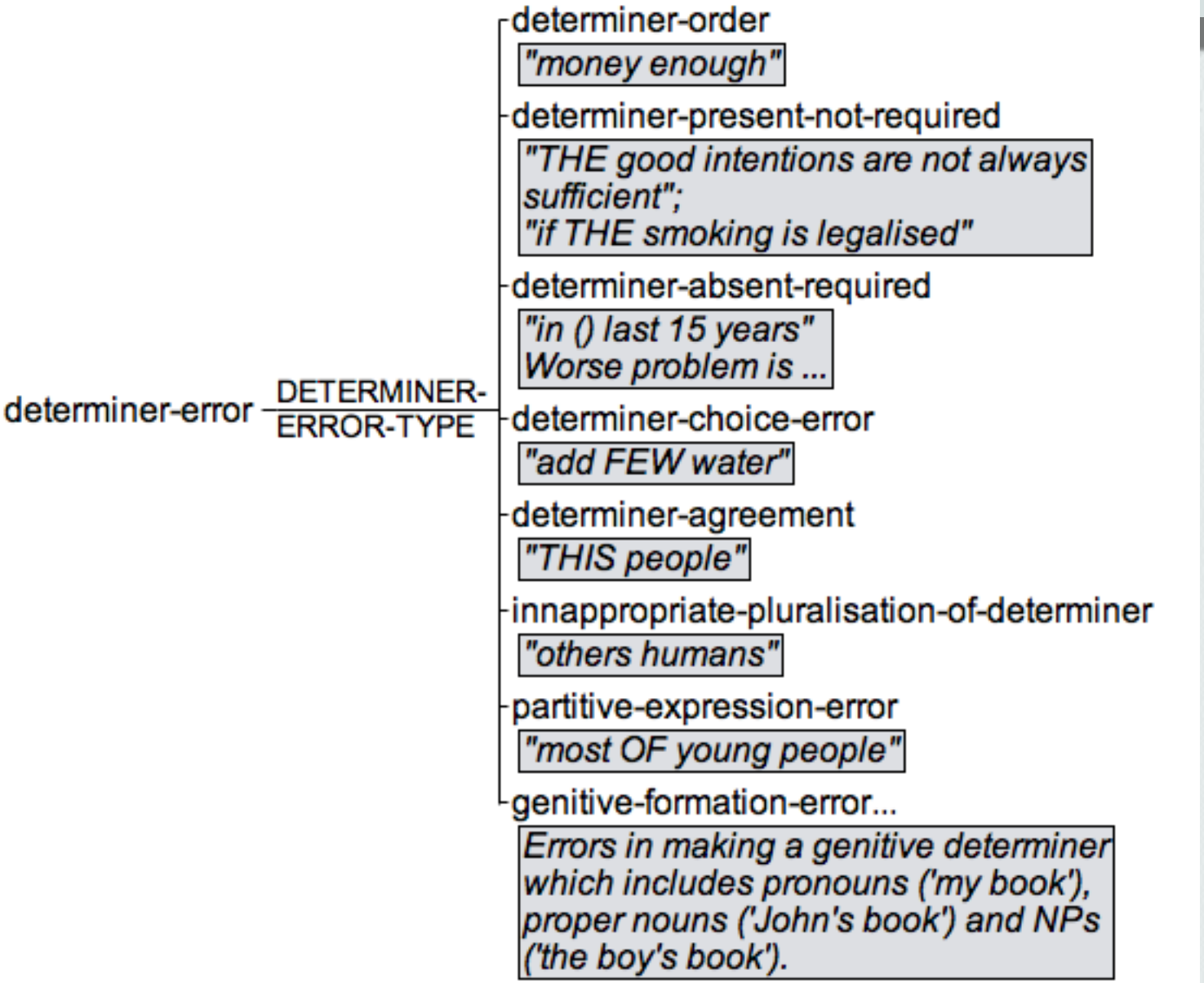
Grammar-error



Grammar-error



Grammar-error



6. Ensuring Inter-Coder Reliability

- We have tried as far as possible to make the coding criteria clear and unambiguous.
- These criteria are available in a 20 page coding criteria document, and also within the coding program itself.
- However, to test how reliably different coders replicate the same results, 7 of us coded 6 new texts (2500 words) for errors, with no discussion between us.
- Around 500 errors in this corpus.

6. Ensuring Inter-Coder Reliability

- Software was written to compare a set of error-coded texts, and produce a “consensus coding”.
- Only includes segments identified by at least 50% of the participants (based on segment bounds only)
- Takes the most common features assigned to each segment.

6. Ensuring Inter-Coder Reliability

- The Consensus included 453 errors

	Mick	Penny	Susana	Keith	Ainoha
Errors recognised	549	540	664	431	604
Segments present in consensus regardless of coding	371	407	422	289	273
%of consensus segments	82%	84%	90%	63%	59%
Segments coded identically to consensus	275	280	328	220	235
%of consensus segments	60%	61%	72%	48%	51%
Segments not in consensus	179	133	242	142	331
Segments in consensus, but coded differently	169	188	129	216	164
Total different from consensus	348	321	371	358	495
% of their segments	63%	59%	56%	83%	82%

6. Ensuring Inter-Coder Reliability

- The Inter-Coder Reliability software produces a document showing each segment identified and how people coded it.

on in Spain is a subject that **given** a lot of play because for one

	Consensus:	grammar-error: vp-error: perfect-formation-error	**gives**
✓	Laura		is
✓	Maria		has given
✘	Ainoha	grammar-error: vp-error: subject-finite-agreement	✓
✘	Mick	grammar-error: vp-error: passive-formation-error	is given
✘	Susana	grammar-error: vp-error: modal-tense-aspect-selection-error	✓
	Comments:	**Mick**: Maybe has been given.	

6. Ensuring Inter-Coder Reliability

Some Comments:

- The levels of agreement are lower than desired.
- However, this was just the first of a planned cycle of three such studies, with each one intended to reveal differences in coding practices, leading to stronger agreement.
- On the basis of this first study, the coding criteria document was revised to cover cases not covered (e.g., segmentation of punctuation marks)
- Also, on analysis of the “Consensus”, often we all agree on an error, but segmented it differently, so disagreement was not real.
- Often real ambiguity as to what the student meant, so hard to decide:

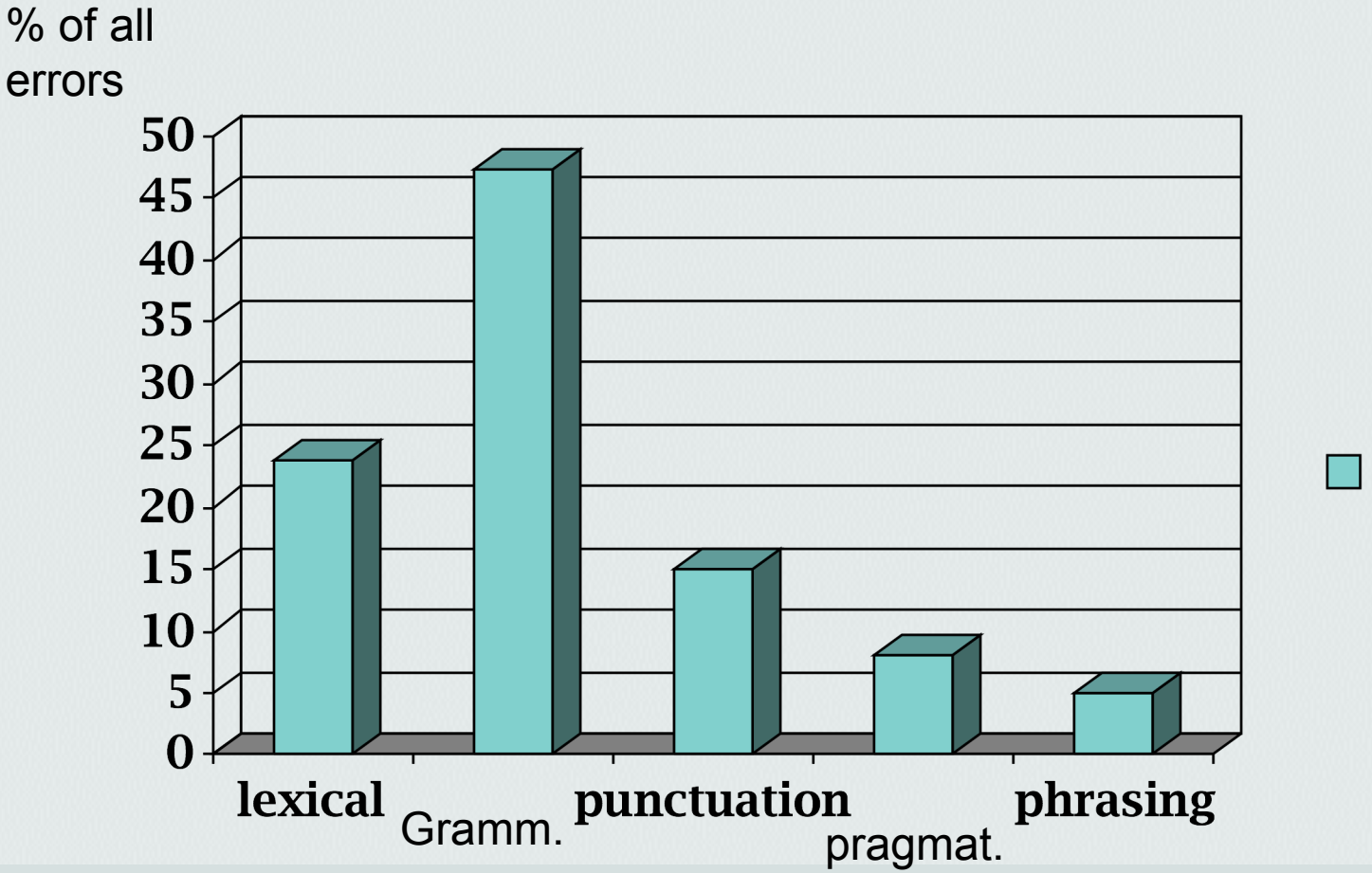
It is shows as we can help

6. Some Results

- Errors coded: 1842
- 28 esays coded, containing 18,400 words
- Just started… …the following results are early, so take them with a grain of salt.

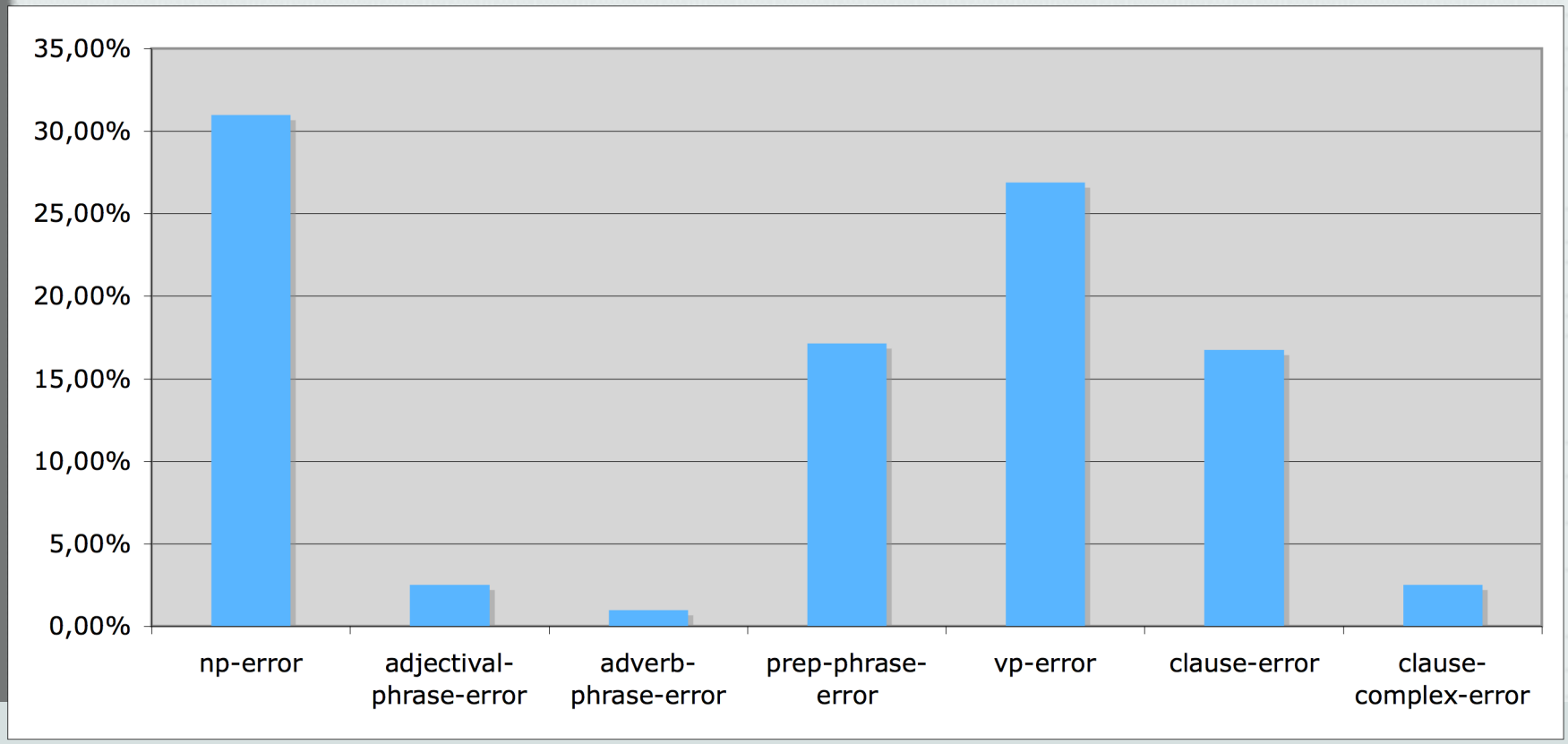
7. Some Results of Our Coding

- By examining the types of errors made at each proficiency level, we can determine how much teaching time to spend on each area.



7. Some Results of Our Coding

- Degree of teaching effort should relate to degree of occurrence of particular errors at the level they are at.
- Looking at graphs per proficiency level provides specific information as to what each group needs
- Use of corpus to derive exercises or examples



8. Future Work

- In parallel work, we are using a parsed corpus to see which where in the proficiency scale syntactic structures start to be used.
- We need to combine this work with the error analysis work.
- 3 stages:
  - Don’t use the structure
  - Use the structure with errors
  - Use the structure correctly

- We are also interested in automatically assigning proficiency level based on the set of errors they make (automatic classifiction using a ‘bag of words’ approach where the ‘words’ are errors)