Ensinar)
Alberto Simões – Universidade do Minho
Diana Santos – Universidade de Oslo
## Ensinador

<table>
<thead>
<tr>
<th>Was designed to...</th>
<th>Was NOT designed to...</th>
</tr>
</thead>
<tbody>
<tr>
<td>To prepare language cloze tests (fill the blanks);</td>
<td>To substitute the teacher;</td>
</tr>
<tr>
<td>To help the teacher;</td>
<td>An automatic tool;</td>
</tr>
<tr>
<td>A semi-automatic tool;</td>
<td>Some tool that someone think might be useful.</td>
</tr>
<tr>
<td>Designed by a teacher using it in the classroom.</td>
<td>A tool for automatic exam correction.</td>
</tr>
</tbody>
</table>
Foreign Language Teaching

- Students should be presented to authentic material;
- With the Internet this becomes trivial;
- Teachers still need to prepare language exercises;
- Preparing exercises based on authentic material is not trivial:
  - Not all authentic material is good for an exercise;
  - Not easy to search for specific language features;
- The use of annotated corpora can help the teacher.
Other Tools

Also used to build cloze tests:
- VISL (Bick, 2005)
- WERTi (Meurers et al, 2010)

Main differences:
- Generate automatic exercises;
- Exercises are not supervised;
- Automatic correction
  - Can discard correct answers;
  - Does not give feedback to the teacher;
Give the power to the Teacher
Support the teacher activity, but do not try to substitute him.
Corpora base

- Ensinador is built over AC/DC corpora;

- AC/DC is a collection of corpora:
  - Different types (newspaper, fiction, interviews...)
  - More than 250 million words
  - Fully annotated using PALAVRAS
  - Being enriched with semantic information (colors, clothes, sentiments)

- AC/DC has been the basis for diverse tools, including concordances in context, frequency service, semantic relation discovery, etc.
Technology base

- Using Open Corpus Workbench;
  - http://cwb.sourceforge.net/
  - Thus, using CQP as base query language.

- Written in Perl;
  - Using CWB Perl modules.

- Interfaces with PHP data files;
  - As AC/DC configuration files are PHP based.

- Use jQuery to facilitate JavaScript interoperability;
  - Minimal amounts of JS, just when strictly necessary.
Functionality

- Start by searching corpora for specific language feature or construct to be exercised;
- Select the best examples to test that feature;
- Search other corpora or refine the search expression;
- Create different exercise sections;
- Download the exercise questionnaire;
- Download the (one possible) correction;
Ensinador Query Syntax

- As stated earlier, based on Corpus Query Processor query language;

- Extended to enable the user to specify annotations to present for each token matching a specific place-holder;

- Each CQP place-holder can be followed by a list of attribute names:

  \[
  \text{[lema="ser"].lema [pos="V"].lema.temcagr}
  \]
Ensinador by Example
Step by step example to create an exercise.
Main Screen

- See current corpus constitution;
- Optionally select another corpus;
- Write a CQP query expression in the “Procurar” Field;
  - [lema="(sempre|nunca)"] [pos="V"]
- Optionally give a name to this exercise section.
  - Nunca/Sempre
Concordance Screen

- Select the lines with interesting examples.
- Remember that in the final exercise, the bold words will be replaced by a line, to the students fill-in;
- It might be interesting to add the lemma of the verb, together with its verbal tense:

  [lema="(sempre | nunca)"] [pos="V"]').lema.temcagr
Concordance Screen

- Note that all examples are now annotated.
- The annotation needs some work (expansion of the annotation codes)
Status Screen

• See your exercise status (as the student will see it)
• Change the lines order (randomize)
• Download the exercise or the solution as an HTML file
• Add a main title to the exercise
• Add more exercise sections;
• Or add more exercises from a different (or same) corpus.
Trans-Ensíñador

- At the moment we are researching on the possibility to expand Ensíñador to parallel corpora;
- Teacher would be able to write query for one language or both languages;
- Probabilistic translation dictionaries (extracted with NATools or GIZA++) could then be used for placeholder alignment.
Concluding Remarks

- Easy to construct real-world examples being selective:
  - The teacher is responsible to pick best examples;

- Make annotated corpora yet more useful:
  - New type of users; make them aware of its relevance;

- More than one correct answer!
  - Without a proper context a sentence can have multiple interpretations;
  - Therefore, automatic correction can not be performed;
  - Interesting to understand students interpretations;
Ensinador is available at
http://www.linguateca.pt/acesso/ensinador