



Fundação para a Computação Científica Nacional



UMIC

Agência para a Sociedade do Conhecimento

MINISTÉRIO DA EDUCAÇÃO E CIÊNCIA



FCT

Fundação para a Ciência e a Tecnologia
MINISTÉRIO DA CIÊNCIA, TECNOLOGIA E ENSINO SUPERIOR



MCTES

MINISTÉRIO DA CIÊNCIA,
TECNOLOGIA E
ENSINO SUPERIOR

Ensinador

Alberto Simões – Universidade do Minho

Diana Santos – Universidade de Oslo

Ensinador

Was designed to...

- To prepare language cloze tests (fill the blanks);
- To help the teacher;
- A semi-automatic tool;
- Designed by a teacher using it in the classroom.

Was NOT designed to...

- To substitute the teacher;
- An automatic tool;
- Some tool that someone think might be useful.
- A tool for automatic exam correction.

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:

`[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 CQPquery expression in the "Procurar" Field;
 - [lema="(sempre | nunca)"] [pos="V"]
- Optionally give a name to this exercise section.
 - Nunca/Sempre

Ensinador

A procurar "[lema="(sempre|nunca)"] [pos="V"]" no corpus Colecção CHAVE v. 4.3 [8000 entradas aleatórias de um total de 21517 entradas.]

Selecione as concordâncias que deseja usar.

- ☐ FSP940101-007: No mesmo clima de oração, poderemos penetrar mais nos desígnios de Deus e descobrir que, até no meio das provações e sofrimentos do ano que passou, há valores que nem **sempre percebemos**.
- ☐ FSP940101-099: Brunner trabalhou na Ferrari de 86 a 88, fez o projeto do carro de Indy da fábrica que **nunca correu** e só saiu de lá quando Barnard chegou pela primeira vez.
- ☐ FSP940101-124: O trabalho de Sims, assim como de Mario Sorrenti e de Corinne Day, trouxe para a fotografia um frescor até então **nunca explorado**.
- ☐ FSP940101-131: O artigo de Otto, à esquerda da página onde **sempre saía**, já dizia no título interrogativo o que buscava: Qual é a fala padrão do brasileiro, agora? "
- ☐ FSP940102-046: Apesar da solenidade da roupa e do tratamento de Vossa Excelência utilizado pelos ministros para se dirigirem uns aos outros, as reuniões plenárias do tribunal têm espaço para humor e para acaloradas discussões jurídicas, que nem **sempre dispensam** fortes toques de ironia.
- ☐ FSP940102-071: Agora, é torcer para que a mesma mudança de comportamento seja adotada por outros formadores de opinião que **sempre estão** na TV, nas rádios ou nos jornais espalhando o pessimismo.
- ☐ FSP940102-071: Estão **sempre inventando** uma desgraça nova.
- ☐ FSP940102-090: Ele **sempre liga** para mim quando tem um freguês hospedado, afirma A.F.S., 21, que adota o nome fictício de Jessica.

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

Ensinador

A procurar "[lema="(sempre)nunca")]" [pos="V"].lema.temcagr" no corpus Coleção CHAVE v. 4.3 [8000 entradas aleatórias de um total de 21517 entradas.]

Selecione as concordâncias que deseja usar.

- ☐ FSP940101-012: Parabéns pelo trabalho de cada um e que Deus continue iluminando o caminho para que a verdade possa **sempre ser** dita. " (, ser INF)
- ☐ FSP940101-131: O artigo de Otto, à esquerda da página onde **sempre saía** , já dizia no título interrogativo o que buscava: Qual é a fala padrão do brasileiro, agora? " (, sair IMPF_IND)
- ☐ FSP940102-008: Uma convicção que deixou aturdidos todos aqueles que **sempre veneraram** Cristo como o exemplo de todos os exemplos . (, venerar PS/MQP_IND)
- ☐ FSP940102-046: Apesar da solenidade da roupa e do tratamento de Vossa Excelência utilizado pelos ministros para se dirigirem uns aos outros, as reuniões plenárias do tribunal têm espaço para humor e para acaloradas discussões jurídicas, que nem **sempre dispensam** fortes toques de ironia . (, dispensar PR_IND)
- ☐ FSP940102-071: Estão **sempre inventando** uma desgraça nova . (, inventar GER)
- ☐ FSP940102-085: Eu **nunca ligo** para dinheiro, a segunda maior cidade do mundo tem que gastar, afirma o secretário de Vias Públicas, Reynaldo de Barros . (, ligar PR_IND)
- ☐ FSP940102-127: A jogadora acredita que o fato de as brasileiras **sempre deixarem** transparecer na fisionomia um ar de abatimento ou derrota visível favorece os adversários . (, deixar INF)
- ☐ FSP940102-133: Lua de Fel entra de sola em algo que **sempre funcionou** com o diretor: sexualidades conturbadas e anexos . (, funcionar PS_IND)
- ☐ FSP940102-151: Essa inserção costuma ser feita com algum vírus inócuo, o que **sempre pode** acarretar o perigo de

Concordance Screen

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

Ensinador [Linguateca](#)

Nunca/Sempre

1. FSP940102-085: Eu _____ para dinheiro, a segunda maior cidade do mundo tem que gastar, afirma o secretário de Vias Públicas, Reynaldo de Barros . (, ligar PR_IND)
2. FSP940107-046: _____ ameaça da Cut ou do PT, disse ele, ao ser indagado se o homicídio não estaria associado às denúncias feitas por seu irmão sobre o uso eleitoral do dinheiro e da máquina do sindicato . (, haver PS_IND)
3. FSP940107-162: Que ela seja _____ quando pensarmos nos pequenos incidentes de nossas vidas . (,

Ver Solução Aleatorizar Descarregar: enunciado solução Adicionar título:

Procurar para a seção no corpus

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-Ensinador

- At the moment we are researching on the possibility to expand Ensinador 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;



Obrigado!

Ensinador is available at

<http://www.linguateca.pt/acesso/ensinador>