

Internet Based Grammar Teaching

Eckhard Bick

Department of Linguistics, Århus University
Institute of Language and Communication, Odense University
e-mail: lineb@ling.hum.aau.dk

Internet Based Grammar Teaching

Eckhard Bick, Århus University, Odense University

1. Abstract

The paper presents an integrated interactive user interface for teaching grammatical analysis through the Internet medium. Though the system's internal grammatical tools for the analysis of free running text are - for reasons of robustness, efficiency and correctness - based on the Constraint Grammar formalism, users are free to choose from a variety of notational filters, supporting different descriptive paradigms. The original kernel of programs was built around a multi-level parser for Portuguese (Bick, 1996 and Bick, 1997) developed in a Ph.D. research framework at Århus University. A similar system has since been implemented for English as part of the VISL-project at Odense University.

2. Introducing and designing IT-based teaching tools

When trying to introduce IT-based tools into a teaching environment, apart from the hardware problem of there never being enough (compatible!) machines at the right place and time, there is the very central problem of psychological resistance against the new medium, simply because it may feel too "technical". Things technical have traditionally a very low acceptance rate in the Humanities, which is where language teaching belongs. Text processors, for example, were widely shunned until the day when they started to use a "non-technical", i.e. graphical, interface. In the same vein, there is the fundamental difference between a human teacher and a computer terminal, - the latter lacks the teacher's naturalness, interactivity, flexibility and tutoring capacities. On the other hand, computers do have evident teaching advantages - they can integrate the senses, making use of colours, pictures and sounds in a more flexible and impressive manner than paper can. Also, a program can "know" more - in terms of facts and data and within a defined specific field - than a human teacher. And last, but not least, a computer program can teach an infinite number of students at the same time in an individual manner, if it is installed on as many machines, or accessible through as many terminals in Internet country.

Given these advantages, it makes sense to invest some effort into addressing the four main disadvantages, named above. My grammar teaching interface tries to make advances with regard to the following four principles:

A) Flexibility

The interface is notationally flexible, i.e. the user can choose one of several notational conventions (e.g. flat dependency grammar, enriched text, meta text notation, tree structures). According to the student's background, the analysis' complexity can be modified, - for example, by increasing or decreasing the number of distinct word class tags used. At the same time, such filtering permits a choice between the use of traditional word class concepts on the one side, or, for instance, purely morphologically motivated ones, on the other side. In order to make a session more colourful, it is also possible to move between corpus text, live newspaper text, randomized test sentences and one's own creative idiolect.

B) Interactivity

A set of CGI-controlled programs reacts instantly to those user choices advocated by the flexibility principle, and the interface changes accordingly in an interactive way, permitting, for example, to move back and forth between levels and notational conventions. When a sentence proves problematic or incomprehensible, the user can modify it, or ask for the computer's opinion. Grammatical analysis can be run in interactive mode, where a sentence is analysed step by step, with the student suggesting form or function readings for words or constituents, and the computer checking and commenting the choices. In this mode the text in question will gradually be coloured for word class and indexed for function as the student's analysis progresses.

C) Naturalness

A major drawback of most language teaching software (or, for that matter, language analysis software) is that they do not run on free, natural language, but on a small set of predefined sentences that cannot be modified or replaced. Usually "toy lexica" and "toy grammars" are used that can only handle a narrow range of built-in structures. In my interface the underlying lexica and grammars cover the whole language, and the user can thus manipulate the text to be analysed in much the same way as in an ordinary text processor.

The second aspect of naturalness concerns, as mentioned above, "untechnicality", and as much keyboard-interaction as possible has therefore been replaced by graphical and mouse based means, like menu choices and clickable radio buttons and help windows. Being internet based, the system automatically takes advantage of a browser's navigation tools, scroll bars, page memory and cut'n'paste functionality.

D) Tutoring

Tutoring is probably that human teacher feature that is hardest to imitate. A teacher's intuitive understanding of a student's problems is difficult to build into a program. A certain minimum of tutoring can be achieved by providing readily available (i.e. "clickable") definitions of grammatical terms, and examples of their usage and the phenomenon's distribution in the language. For the latter purpose, a powerful corpus searching tool has been crafted to find examples of user-defined grammatical structures in automatically (and, in theory, simultaneously) annotated corpora at the system's disposal. After acquiring some basic notational skills a student (or researcher) can search for any combination and sequence of word forms, lexemes, word classes, syntactic function and so on. Ultimately, "guided tours" could be designed for certain topics by blending the definition and corpus example tools.

Another aspect of tutoring, that proves useful in *foreign* language teaching, and has been tentatively introduced for Portuguese, is translational help, either in the form of dictionary enquiries, translational tagging¹ or even rough sentence translation.

¹ Here, base form translation equivalents are given as the last tag on the tag line in verticalized CG-notation. Some polysemy resolution is performed, based on valency instantiation and the disambiguation of atomic semantical features by Constraint Grammar rules.

3. The grammatical base

The grammatical backbone of both the Portuguese and the English system is a combination of lexicographical data bases with disambiguation and mapping grammars in a Constraint Grammar frame work (Karlsson, 1995). Large base form lexica with inflectional, word class and valency information are used by a morphological tagger, whose output ambiguity is then resolved with the help of Constraint Grammar rules², that "prohibit" certain readings in certain sentence contexts. Syntactic function labels are then mapped onto the morphologically disambiguated word forms, and a second round of disambiguation rules tackles the new - syntactic and much larger - ambiguity, drawing on global sentence context, word class sequences, agreement patterns, valency class information and the like. In principle, this process can be iterated on ever higher levels (Bick, 1997 and forthcoming), and is, in fact, used in the Portuguese system for helping the student by providing disambiguated base form translations or even crude MT on running text.

But why Constraint Grammar?

* Constraint Grammar is robust. A language teaching system based on natural text must be very stable, and be able to provide *some* analysis to all input. A "no parse"-message window would destroy the illusion of a real teacher, and - if frequent - ultimately result in student frustration. Since CG works by adding and removing information, the *correct* reading will crystallize in an indirect way - simply by being the last surviving analysis. Thus, in the CG formalism, even unusual or partial sentences will receive *some* analysis, and an ill-formed sentence will not prevent correct *lower* level analysis, for instance, correct word-class and noun phrase analyses.

* Constraint Grammar is tag based, and adds tag strings to word forms. First, string based information is easy to port and easy to manipulate in a computer, and second, this way different kinds of information, lexical, morphological and syntactical, form, function and structure, can be handled withing the same formalism³, which allows easy notational transformation. Thus, tags can be fused into more general Portmanteau-tags (downward compatibility), og split up into subcategories by using higher lever information from other tags in the same string (upward compatibility). An example for the first is the fusion of adverbial adjuncts, adverbial objects and prepositional objects into a Portmanteau-tag 'adverbial', and an example for the latter is the function-based distinction between "adjectival" (adjective-like) and "substantival" (noun-like) pronouns. With sufficiently

² In the English system, a commercial CG-parser (© lingsoft, Helsinki, Finland) is used as a base (for a discussion of the parser see [Voutilainen, 1992]), and additional programs are used to adapt and integrate the parser's output annotation. Finally, an additional layer of CG-rules is applied in order to reduce ambiguity, identify subclauses and tag for subclause function. The Portuguese system has been developed by myself over a 5-year period, much of it in a Ph.D. research framework.

³ Of course, this is true of the *internal working* of the grammar, too. Constraints can be worded in much the same way whether they are morphological, syntactic or semantic, and information from different levels can interact in disambiguation.

detailed dependency markers, CG-notation can even be transformed into constituent based tree structure notation (Bick, 1997-1).

* CG-notation has elegant ways of underspecifying ambiguity. Postnominal PP-attachment, for example, is expressed as "nominal attachment to the left" (@N<), so that the Chinese origin in *The man with the bicycle from China* can be applied to both 'man' and 'bicycle'. In cases of ambiguous functional analysis, CG can add several (competing) function tags to the same word, so that the ambiguity can be expressed in *one* analysis. Especially with long sentences this is paedagogically superior to having to scroll through several pages with tens or hundreds of possible analyses. Also, it becomes easier to judge the student's analysis - if the tag suggested is a substring of the ambiguous tag string, then the suggested reading will be accepted by the computer, even if it is not the only one. For the same reason, if the computer fails to resolve some ambiguity, this will not impair the student-computer interaction, - as long as the correct reading is among the ones "surviving" the CG-treatment (which can be geared to prefer ambiguity to errors), the robot teacher may be over-indulgent, but it will not harshly criticize a justifiable student choice.

* The parser supplying the CG analysis to be used by the interface, is both modular and incremental in its structure. Due to its modularity, for a growing system, one can choose those modules that already have achieved a sufficiently high level of correctness and coverage, and make them accessible to the student community. The English system, for example, does not cover subclause function, but it is still fully operational, within the same interface, on the word class and phrase levels. The CG-parser's incrementality lets the system grow like a holographic picture - the object is visible all the time, only its granularity improves with the amount of time and work put into it. Once the user interface is in place as such (and hardware and wiring technology permits), there is a teaching and demonstration dividend, even if the parser can still be improved. Thus, with a CG-parser, the time lapse between grammatical work and paedagogical implementability can be reduced to a minimum.

4. The paedagogical base

Word based tags (after, under, over , indexed or - as colour code - "inside" the words in question, with or without underlining, in the form of abbreviations or symbols) are paedagogically intuitive and close to "basic" grammar, - not only for marking word class, but also in syntax, as can be told from the "cross-and-circle" grammar used in Danish primary schools, or the corresponding colour-underlining system used in Germany. A special advantage of CG's dependency notation is that it mirrors children's semantically based intuition making the head of a phrase the bearer of its syntactical weight. For the sentence "*Pia's stupid rabbit ate the flowers I collected for mother*" the quick answer to the subject question ("*Who ate ...?*") is "*The rabbit!*" and, even more surely, the answer to the object question ("*What did the rabbit eat?*") will be "*The flowers*". It usually takes additional syntactic curiosity from the teacher's part to elicit answers as to whose rabbit and which flowers the sentence was about. Apart from articles (that are necessary to state a noun's definiteness, something which can be achieved in Danish by morphological means),

most other modifiers seem to be outside the reach of "subject"-ness or "object"-ness. Most strikingly so in the case of parenthetical relative clauses: *Ann, who hadn't slept for two nights, wanted to go home - Who wanted to go home? - Ann*. Here, dependency analysis seems to be mentally more basic than constituent analysis, which becomes secondary: It is the subjects 'rabbit' and 'Ann' that grow into complex constituents by absorbing modifiers like 'stupid' or 'who hadn't slept for two nights', - and not a subject constituent that breaks down into several sub-constituents. I believe, that it is pedagogically important to start from the (concrete) referent center (i.e. 'the rabbit' and 'Ann') and work from there by adding more and more bricks (each of them still as small and as concrete as possible), creating a - larger - whole that is still concrete in the child's mind, in stead of starting with an abstract unit (a subject constituent) that will not be made concrete but several layers of analysis further down (i.e. at word level).

In my system's interactive grammar module, this thought is matched both by notation and procedural sequentiality: Functions are tagged at a phrase's head word, and it is possible to correctly click and identify, say, a subject head as "subject", even *before* possible modifiers have been attached by additional clicks and menu-choices. This contrasts strongly with a traditional constituent based approach, where there can be no subject without a subject *constituent*.

Still, while advocating a head-driven and bottom-up analysis for pedagogical and psycholinguistic reasons, the flexibility principle is applied to this matter, too, and students do have the choice of a tree-structured constituent analysis (which, for Portuguese, is automatically derived from the flat dependency notation), thus facilitating a top-down perspective where desired⁴.

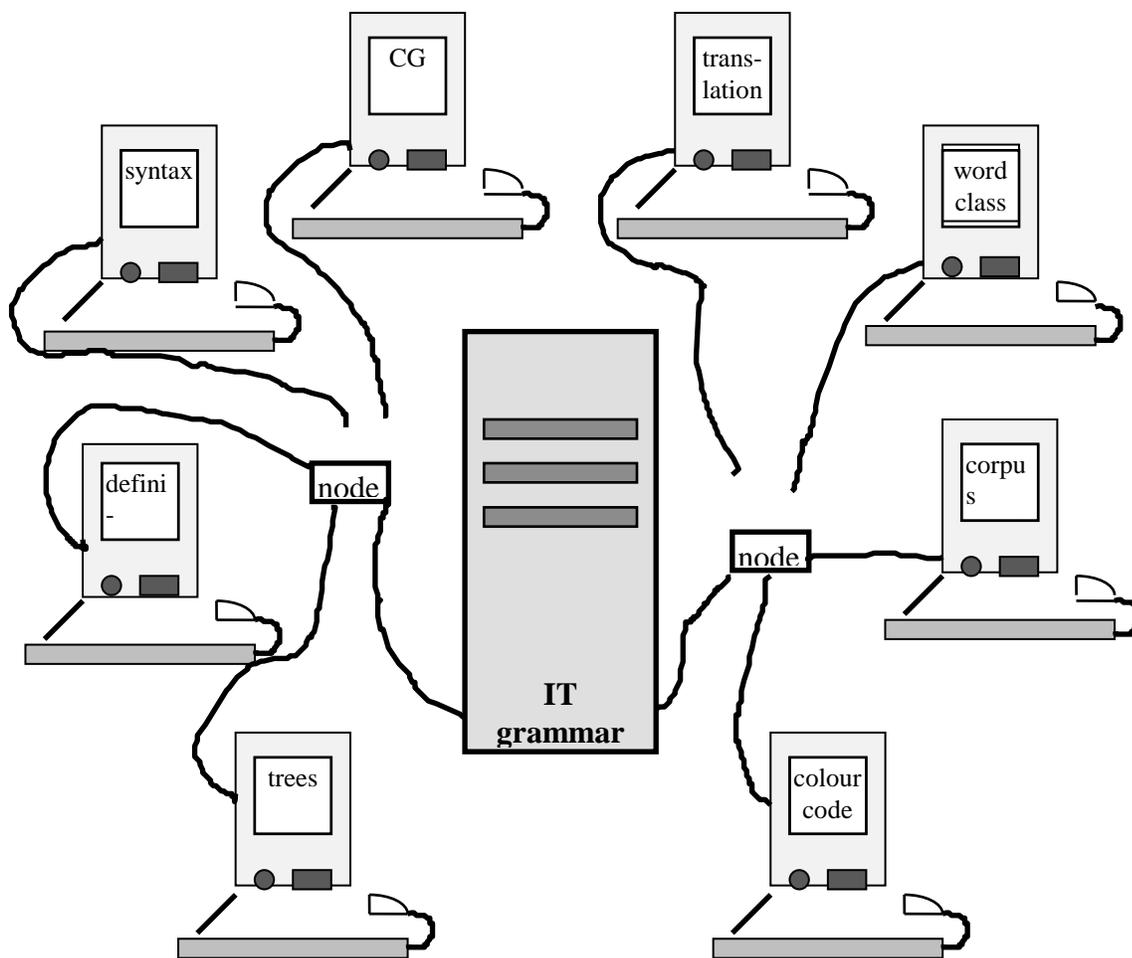
5. The interface

The system is implemented as a free-for-all distributed teaching environment, with one or more servers running the grammar software and the CGI-programs necessary to interact through the internet with users at their school, university or home computers (1).

(1) Distributed grammar teaching environment: a central IT-grammar server handles - in parallel - a large number of student terminals that may focus on different languages, different levels of analysis or different training tasks, representing different notational or grammatical systems.

⁴ For now, this notation does not yet permit *interactive*, i.e. student driven, analysis the same way the dependency grammar notation does, but one of the spin-offs of the VISL-project at OU will be java software designed to address this problem, as well as - hopefully - a more flexible CG-compiler capable of permitting context sensitive notational modification on top of a pre-existing CG-grammar (like tag replacement and the introduction and removal of explicit constituent boundaries).

A first version of a tree-drawing java program was written by Thomas Larsen and is now being modified and extended by Martin Carlsen for the VISL-project at Odense University.



With the exception of a tree-drawing java program⁵ for the English module, all computation is done *at server side*, and user input and choices are managed through server-updated html-forms. This method has a number of advantages over traditional *user side* based software: First, no software has to be sold or distributed and, consequently, copy-right infringement is minimized; second, the age and quality of the user's computer is of less importance (as long as it can run a browser), and - not entirely unimportant for multi-language applications - incompatibilities with regard to software, character set, machine type etc. are circumvented; third, interaction is speedy, since only short html-texts are sent back and forth, while programs proper are run by those machines that are good at it - heavy computation intensive grammar programs by the server, light keyboard, mouse and text manipulation by the terminal machine whose language and other preferences remain customized by the user.

The flow chart diagram (2) illustrates the interaction between student and grammar server in a sequential way, pointing out where and how information is provided and what choices can be made by the student in order to navigate through the teaching system.

⁵ cp. footnote (4)

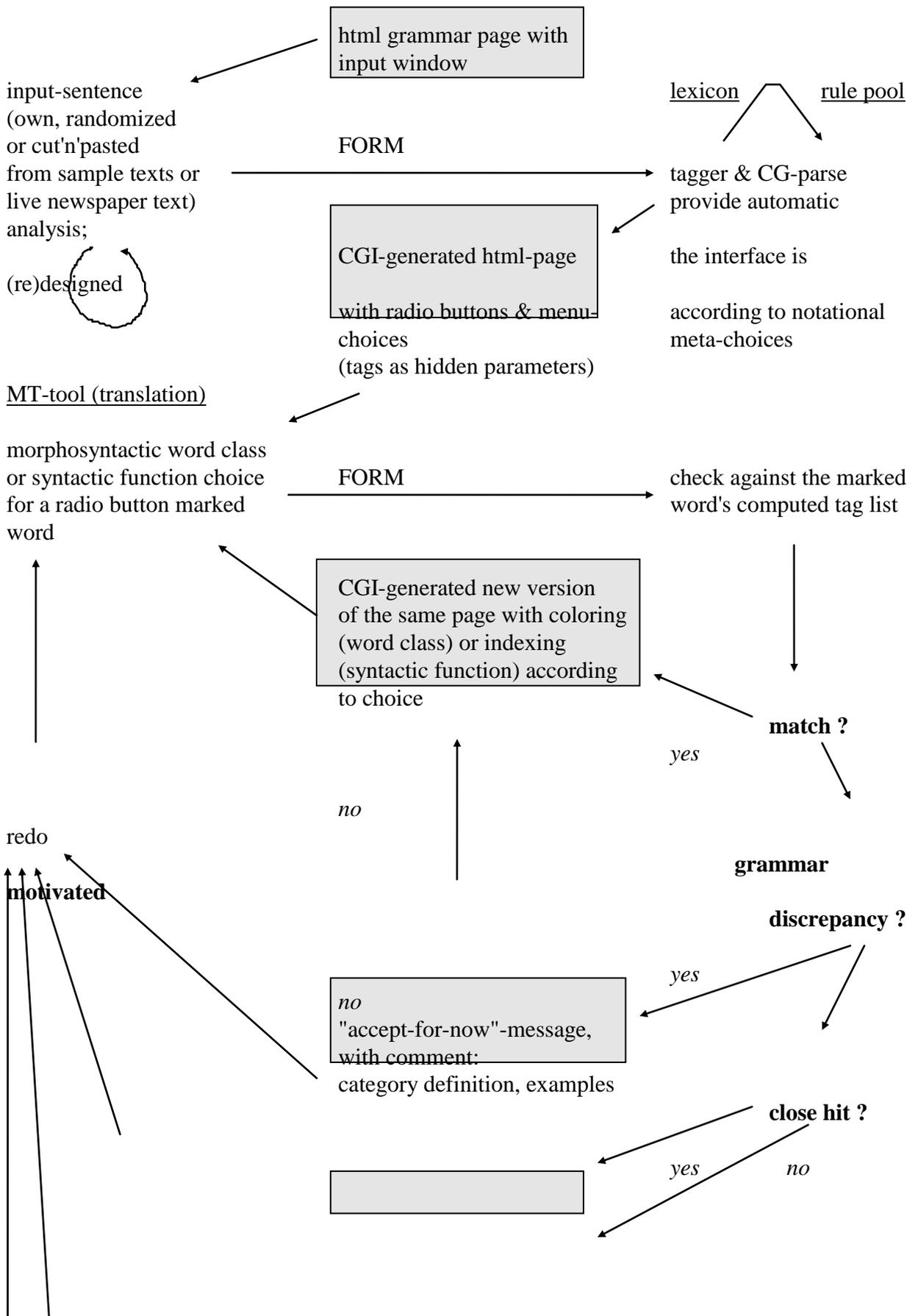
(2) flow chart of student - server interaction:

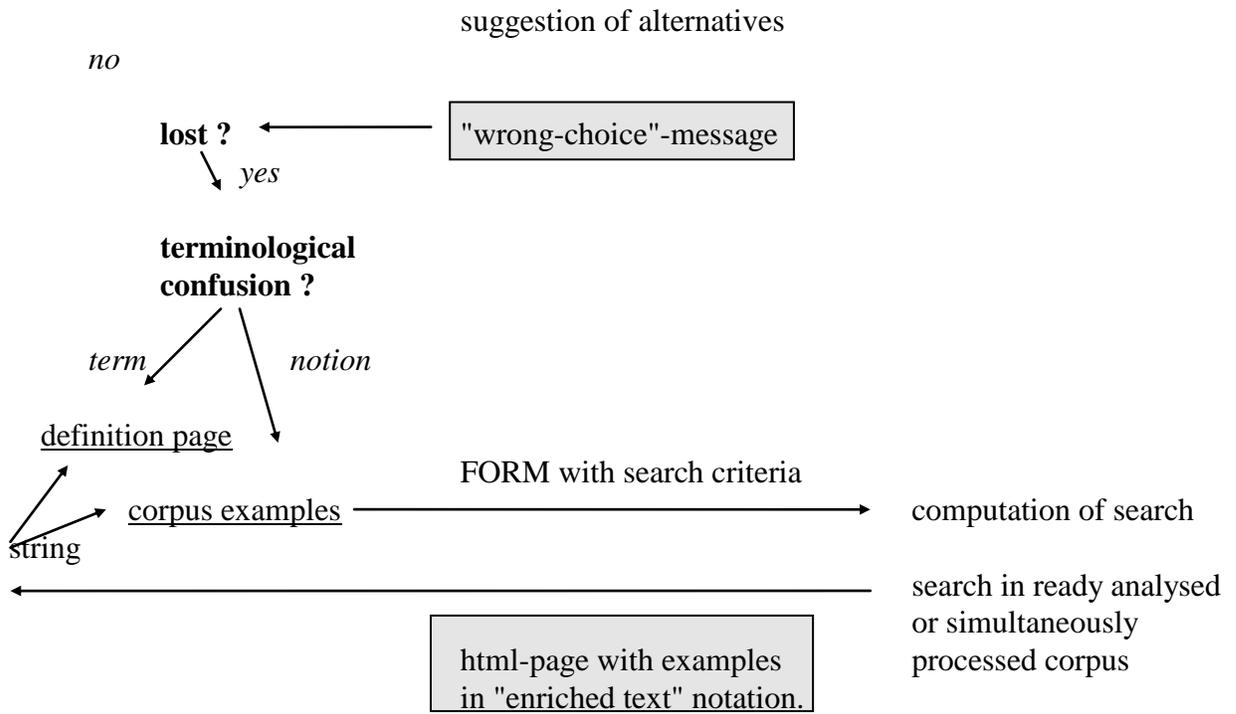
CLIENT SIDE

(Student's terminal computer)

SERVER SIDE

(Grammar host computer)





6. A guided tour

The Portuguese module runs at "http://ling.hum.aau.dk/~eckhard/Linguistics.html" and "http://visl.hum.ou.dk/Linguistics.html", presenting the following web-page:

(3) The Portuguese grammar page

Netscape: Eckhard Bick: Portuguese Linguistics

Automatic Analysis of Portuguese

If you want to test the Portuguese parser and the VISL grammar interface, you can do so here. Press a radio button for mode of analysis, and enter your text in the window below. If you don't know any Portuguese, but still want to test the parser, you can copy & paste some sentences from [a Portuguese novel](#) or [Brazilian newspaper text](#).

Portmorf Portpars Porttrad Portdan Flatmorf Flatsyn

Interactive grammar, using

Dictionary (Portuguese or Danish) Corpus search (sequence of words/tags) V-trees H-trees

Enter new text for analysis (default example given), translation or search:
Empty window in interactive grammar mode yields randomized sentences.

Ficar sem trabalho é ruim para qualquer pessoa, mas no caso de um executivo a demissão vem acompanhada de uma série de mudanças que muitas vezes acabam comprometendo a própria chance de conseguir uma nova colocação.

[Reset](#)

Task Definitions:

- Portmorf - all morphological possibilities
- Portpars - disambiguated morphosyntactic analysis
- Porttrad - disambiguated morphosyntactic analysis and base form translation
- Portdan - translation into Danish
- Flatmorf - running word class color notation
- Flatsyn - running word class color notation with syntactical indexing
- Interactive grammar - build your own analysis (form and function), assigning menu item categories to each word in turn, with the computer tutoring your choices
- Dictionary search - presents layouted dictionary entries for base forms of words og expressions, works both Portuguese - Danish and Danish - Portuguese
- Corpus search - search for words, base forms, grammatical category etc. , or combinations of these, in a corpus of written Portuguese (word forms in double quotes, base forms in single quotes, tags separated by blank space, words by underscore, use '_' for facultative dummy word, ' ' [space] or '' [nothing] for obligatory dummy words)
- V-trees - vertical tree analysis (enriched Constraint Grammar)
- H-trees - horizontal tree analysis (traditional)

The page asks the user to enter Portuguese text, and less inventive souls (or curious people without any deeper knowledge of Portuguese) are offered a default example as well as sample text or newspaper links for cutting and pasting. Next, there is a choice between different levels of analysis, from simple tagging ('Portmorf') over morphosyntactic disambiguation ('Portpars') to bilingually motivated polysemy resolution ('Porttrad'),

between different notational conventions (verticalized word based CG notation, enriched text ['flatmorf' and 'flatsyn'] with meta-tagging as well as tree-structures ['V-trees' and 'H-trees']) and - in interactive emode - between two levels of descriptonal complexity ('full tag set' and 'traditional tag set').

In "passive" mode where interactivity does not concern grammatical analysis proper but only text and descriptonal choices, the system functions much like a multi-level, notationally flexible parser. Thus, a "raw" Constraint Grammar analysis, with full disambiguation, yields the following output for the default sentence given:

(4) Full multi-level Constraint Grammar analysis

disambiguated morphosyntactic analysis and base form translation

Ficar sem trabalho é ruim para qualquer pessoa, mas no caso de um executivo a demissão vem acompanhada de uma série de mudanças que muitas vezes acabam comprometendo a própria chance de conseguir uma nova colocação.

Ficar [ficar] <vK> <v-cog> **V INF 0/1/3S @IMV @#ICL-SUBJ> 'blive'**
sem [sem] **PRP @<SC 'uden'**
trabalho [trabalho] <d> <am> **N M S @P< 'job'**
é [ser] <vK> **V PR 3S IND VFIN @FMV 'være'**
ruim [ruim] <+para> <jh> <jn> **ADJ M/F S @<SC 'dårlig /for'**
para [para] <+hum> <move+> **PRP @A< 'til'**
qualquer [qualquer] <quant2> **DET M/F S @>N 'en hvilken som helst'**
pessoa [pessoa] <H> **N F S @P< 'person'**

,
mas [mas] **KC @CO 'men'**
no=caso=de [no=caso=de] <c> **PRP @ADVL> 'i tilfælde af'**
um [um] <quant2> <arti> **DET M S @>N 'en'**
executivo [executivo] <prof> <HH> **N M S @P< 'funktionær'**
a [a] <art> **DET F S @>N 'den'**
demissão [demissão] **N F S @SUBJ> 'demission'**
vem [vir] <vi> <sN> **V PR 3S IND VFIN @FMV 'komme'**
acompanhada [acompanhar] <vH> **V PCP F S @<PRED 'ledsage'**
de [de] **PRP @A< @<ADVL 'af'**
uma [um] <quant2> <arti> **DET F S @>N 'en'**
série [série] **N F S @P< 'følge'**
de [de] **PRP @N< 'af'**
mudanças [mudança] <cP> **N F P @P< 'forandring'**
que [que] <rel> **SPEC M/F S/P @SUBJ> @#FS-N< 'som'**
muitas=vezes [muitas=vezes] **ADV @ADVL> 'ofte'**
acabam [acabar] <x+GER> **V PR 3P IND VFIN @FAUX 'ende /med at'**
comprometendo [comprometer] <vt> <vH> **V GER @IMV @#ICL-AUX< 'kompromittere'**
a [a] <art> **DET F S @>N 'den'**
própria [próprio] <ident> **DET F S @>N 'selve'**
chance [chance] <ac> <+de+INF> **N F S @<ACC 'chance /for'**
de [de] **PRP @N< 'af'**
conseguir [conseguir] <vt> <vH> **V INF 0/1/3S @IMV @#ICL-P< 'opnå'**
uma [um] <quant2> <arti> **DET F S @>N 'en'**
nova [novo] <ante-attr> **ADJ F S @>N 'ny'**
colocação [colocação] **N F S @<ACC 'opstilling'**

.

Output convention: (click for tag list and definitions):

text word [base form] <derivation, valency etc.> **CATEGORY INFLECTION @SYNTAX stem translation**

Here, a sentence' word forms (bold face brown) are verticalized and their concerning tags coloured for lexeme (light brown), word class (bold face blue), inflection (light blue), clause internal syntactic function (light green), clause function (bold face green) and base form translation (yellow). The angle bracketed tags provide additional "secondary" information (which has also been partially disambiguated) about, for instance, subclasses like 'relative', 'interrogative', 'demonstrative' and so on for pronouns, as well as valency patterns used in the context given. By following the concerning links at the bottom of the page the student can find help with regard to category definitions, abbreviations and the like. Contentwise, he can deduce a rough translation from the Danish equivalents offered as part of the tag string, or he may ask for additional help in the form of a "real", running translation of the sentence ('Portdan'):

(5) live, CG based, machine translation

translation into Danish

Ficar sem trabalho é ruim para qualquer pessoa, mas no caso de um executivo a demissão vem acompanhada de uma série de mudanças que muitas vezes acabam comprometendo a própria chance de conseguir uma nova colocação.

at blive uden job er dårlig for en hvilken som helst person, men i tilfælde af en funktionær kommer demissionen ledsaget af en følge af forandringer som ofte ender med at kompromittere selve chancen for at opnå en ny opstilling.

Output convention: (click for tag list and definitions):

[Home Page](#) (Eckhard Bick)
[Questions and comments](#) (e-mail)

Though the system's MT is often fairly crude for longer sentences, it does, for instance, handle NP-agreement, basic word order rules and some prepositional valency and incorporation for verbs.

Output as in (4) is close to the grammatical core of the system and combines most of the CG advantages listed in chapter 3. For many purposes, however, this very detailed notation may seem too heavy a tool, especially if the user has no prior experience with Constraint Grammar. According to the principle of naturalness, one would prefer a notation as close to ordinary text as possible. That way, sentence context will be easier to grasp, and the interface will feel less "technical" (as intended). I believe to have found such a notation in what I call "enriched text", where running text is "meta-tagged":

(6) "Enriched text" (running text with meta-tagging)

running word class color notation with syntactical indexing

Ficar sem trabalho é ruim para qualquer pessoa, mas no caso de um executivo a demissão vem acompanhada de uma série de mudanças que muitas vezes acabam comprometendo a própria chance de conseguir uma nova colocação.

Ficar ^{ICL-SUBJ>}_{IMV} **s**em ^{<SC} **t**rabalho **é** ^{FMV} **r**uim ^{<SC} **p**ara **q**ualquer **p**essoa **, mas** ^{CO}
no=caso=de ^{ADVL>} **u**m **e**xecutivo **a** **d**emissão **v**em **a**companhada **de**
^{<ADVL} **u**ma **s**érie **d**e **m**udanças **q**ue **m**uitas=vezes **a**cabam **co**
mprometendo ^{ICL-AUX<} **a** **p**rópria **c**hance **d**e **c**onseguir **u**ma **n**ova **c**olocação ^{<ACC}

Output convention: (click for tag list and definitions):

[WORD CLASS DEFINITIONS](#)

[SYNTACTICAL CATEGORY DEFINITIONS](#)

noun **N**, proper noun **PROP**

personal pronoun **PERS**, "nominal" pronoun **SPEC**, determiner pronoun **DET**

adjective **ADJ**, adnominal participle **PCP**, numeral **NUM**

verb **V**, verbal participle **PCP**

adverb **ADV**, preposition **PRP**, conjunction **K/S/KC**, interjection **IN**, affix **EC**

[Home Page](#) (Eckhard Bick)

Here, each line of the CG-notation is condensed into its text kernel, the word form as such, which is all that is left *on* the line. Thus, the impression of running text is recreated. Of the original tags, only syntactic function is retained, with clause internal function as sub-scripts, and clausal function as super-scripts. Word class is retained as meta-notation, too, in the form of colour codes⁶ (which are explained at the bottom of the page). Nominal material is tagged in different shades of blue so as to retain NP-coherence in a visual, paedagogically intuitive way. Thus, nouns are blue, proper nouns dark violet and adjectives green. Pronouns match what they are pro-forms for - personal pronouns are coloured light violet, nominal (non-inflecting) pronouns turquoise, and determiner (inflecting) pronouns olive-green. With a grass-green shade, numerals belong in the nominal modifier (adjective) camp, too. Verbs receive an entirely different ("active") colour, red, so as to make them stick out from the rest of the sentence. Since also infinitives and gerunds are coloured red, the whole verb chain is easily detected. Participles, being a morphological class capable of both "verbal" and "adjectival" function, are tagged according to syntactic function - as part of a verb chain they are red, but in adnominal position they become as green as ordinary adjectives. The non-inflecting particle classes, finally, divide the remaining colours among themselves, - adverbs, for instance, are yellow, and prepositions brown.

While the "enriched text" notation is ideal for combining the natural flow of running text with word class and function information in a graphical way, it does not emphasize constituent structure. Rather, the latter is expressed indirectly, and in a flat way, by collecting heads and dependents into constituents with the help of directed "dependency markers": @>N, for example, points to a nominal head to the right, thus signalling a prenominal modifier, while @N< stands for a postnominal modifier. Still, the dependency

⁶ With regard to the colour notation of word classes categories I have been inspired by a similar notation, Gratex, for pre-analysed Danish text, described in (Lytje & Donner, 1996).

(8) The English grammar interface⁷ (at "http://visl.hum.ou.dk/engelsk.html")

The screenshot shows a Netscape browser window titled "Netscape: English VISL". The page content is as follows:

English Grammar Page

This is the laboratory version of a VISL-style English Grammar page. You can either (A) work with free text, or (B) choose and clip a sentence from [our closed corpus](#). You may also run a corpus search in order to find a sentence containing a certain word chain, grammatical feature or syntactic pattern.

A) CG-based Flat Dependency Grammar interface (unrestricted input)

CG-Parse Flat syntax Open corpus search

Interactive grammar, choose learning level:

B) Syntactical tree structures for constituent analysis (closed corpus only)

Closed corpus search (sequence of words/tags) V-trees H-trees S-trees

Enter new text for analysis or search:
Empty window in interactive grammar mode yields randomized sentences.

[Empty text input area with scrollbars]

[Navigation buttons: back, forward, home, stop]

[Analysis] [Reset]

7. Student driven interactive analysis

In the interactive analysis mode, a full analysis is computed by the server, but the Constraint Grammar tags remain concealed as hidden parameters in the html-forms sent back and forth through the CGI-channel. Text is presented as running word forms with "clickable" radio buttons, and tag options are presented as menu choices:

(9) text with radio buttons and tag-menus for progressive interactive analysis

⁷ The page allows navigation to a VISL home page and offers links to explanatory pages on the grammatical concepts and conventions used. It also provides links to corresponding VISL grammar pages for other languages, including - apart from Portuguese - German ("http://visl.hum.ou.dk/tysk.html") and French ("http://visl.hum.ou.dk/fransk.html"), with Spanish, Danish and Arabic in waiting.

Since the English interface's analysis of *open text* (A) is based on a CG that is still being modified, users with a special interest in CG are allowed to compare "raw tagging" with different levels of syntactic disambiguation (heuristics or not, subclause level or not), by choosing from a special CG-parse menu.

Select a word, and chose the appropriate word class and/or function! Group function is represented at the head word, subclause function at the main verb (infinite clauses) or the complementizer (finite and verbless clauses).

- Ficar sem trabalho é ruim para qualquer pessoa , mas no=caso=de
 um executivo a demissão vem acompanhada de uma série de
 mudanças que muitas=vezes acabam comprometendo a própria chance de
 conseguir uma nova colocação .

Morphology	Word or group function	Clausal function
noun	main verb	none
proper noun	auxiliary	argument of auxiliary
adjective	subject	subject
adverb	direct object	direct object
personal pronoun	dative object	subject complement
determiner pronoun	subject complement	object complement
non-inflecting pronoun	object complement	prepositional object
finite verb	prepositional object	adverbial object
participle	adverbial object	adverbial
infinitive	adverbial	free nominal adjunct

[New text](#) (back to Linguistics page)

[Home Page](#) (Eckhard Bick)
[Questions and comments](#) (e-mail)

The first menu is primarily about word class, but makes - in addition - a morphological distinction between 3 types of non-inflecting verb forms (infinitives, participles and gerunds). The second menu selects word or group function, with the latter to be marked on the group's head word. The last menu, finally, allows to add subclause function, which is assigned to main verbs in non-finite subclauses, and to complementizers (conjunctions, relatives, interrogatives) in finite or verb-less subclauses. Since complementizers are obligatory for these clause types in Portuguese, but not in English, this convention has been changed in the English module, and subclause function is here always tagged on the clause's first verb, whether finite or not.

In the example, the student has chosen to analyse the third word of the sentence, "trabalho". He can now chose a form or function tag - or both - from the menus. If he - correctly - choses 'noun', the word in question will be coloured blue, and he is allowed to go on with another tag. One of the most simple exercises, which can be carried out even by primary school children, would be to identify, say, *all* the nouns, with correct choices backed by progressive colouring of the sentence, as seen in (10). Note, that the last noun's radio button has disappeared, since it has also been tagged for function (here, direct object). For the leading verb (in the infinitive) full analysis means *two* function tags, since the student here also needs to identify the subject function of the infinitive-clause as a whole (shown as super-script). This way, the sentence appearance will gradually change into the "enriched text" notation (6), to which it becomes identical after full correct analysis:

(10) intermediate stage interactive analysis

morphology missing ...
word/group function missing ...
clause function was ok.

Ficar ^{ICL-SUBJ} _{INV} sem **trabalho** é ruim para qualquer **pessoa** , mas
 no=caso=de um **executivo** a **demissão** vem acompanhada de uma
 série de **mudanças** que ^{FS-H<} muitas=vezes acabam comprometendo a
 própria **chance** de conseguir uma nova **colocação** _{<ACC} .

Morphology	Word or group function	Clausal function
noun	main verb	none
proper noun	auxiliary	argument of auxiliary
adjective	subject	subject
adverb	direct object	direct object
personal pronoun	dative object	subject complement
determiner pronoun	subject complement	object complement
non-inflecting pronoun	object complement	prepositional object
finite verb	prepositional object	adverbial object
participle	adverbial object	adverbial
infinitive	adverbial	free nominal adjunct

[New text](#) (back to Linguistics page)

[WORD CLASS DEFINITIONS](#) [SYNTACTICAL CATEGORY DEFINITIONS](#)

noun **N**, proper noun **PROP**
personal pronoun **PERS**, "nominal" pronoun **SPEC**, determiner pronoun **DET**
adjective **ADJ**, adnominal participle **PCP**, numeral **NUM**
verb **V**, verbal participle **PCP**
adverb **ADV**, preposition **PRP**, conjunction **KS/KC**, interjection **IN**, affix **EC**

[Home Page](#) (Eckhard Bick)
[Questions and comments](#) (e-mail)

In the example, the student's last choice has been a (postnominal) subclause function label for the relative pronoun "que". Since my grammar - somewhat unorthodoxly - defines 3 pronoun classes in a purely morphological manner by lexeme and word form categories, he may have found it difficult to decide on one or other pronoun class. He can now continue in one of three ways:

(a) he can chose (maybe at random) a pronoun subclass and wait for comments. In the case of a wrong choice, the system will act teacher, accepting his choice for being within the pronoun class (which must be duely honoured) while at the same time explaining why the system prefers another category, and how this category is defined. Here, the paedagogical strategy is to distinguish between "absolute errors" and errors originating from the clash of two conceptually different schools of grammar. In the latter case, - be the teacher human or not -, the student's view should be accepted for what it is, and the difference be explained.

(b) he can scroll to the bottom of the menu window and select the last item, "Show me!". The system will then show the correct analysis and colour/index the word in question accordingly. Especially for the word or group function menu it proved unavoidable to introduce this choice, due to the highly differentiated tag set used - and, of course, so as not to frustrate the student unnecessarily. Also, since "live" text is being used, there *is* a chance - though a tolerably small one - for the system being wrong, and the student's analysis right.

(c) finally, there is the possibility of switching to a more traditional tag set by means of a special meta-menu among the navigation buttons underneath (now showing "full tag set" mode). The tag-menus will then be simplified, and there will be only one pronoun class, with articles forming a new, independent class. Similarly, for function, "adverbial object" (i.e. valency bound adverbial) and "adverbial adjunct" (or even "prepositional object") will be fused into the Portmanteau tag "adverbial".

If an error is made, even if it does not originate from a different view on the categories of grammar, it may still be a "soft" error, where the student is fairly close to the correct answer. In (11), the participle "acompanhada" has been called for a gerund. Since both categories are clearly verbal (and even, both non-finite), the system does not simply reject the answer as "plain wrong", but accepts the "verbality" as correct and encourages further subdivision:

(11) Tutoring in the case of a "close miss"

It does belong to the verbal class, but the morphology is wrong. Try another verbal category.
word/group function missing ...

Ficar INV ICL-SUBJ sem **trabalho** é ruim para qualquer **pessoa** , mas
 no=caso=de um **executivo** a **demissão** vem acompanhada de uma
 série de **mudanças** que FS-N< muitas=vezes acabam comprometendo a
 própria **chance** de conseguir uma nova **colocação** <ACC .

Morphology	Word or group function	Clausal function
noun	main verb	none
proper noun	auxiliary	argument of auxiliary
adjective	subject	subject
adverb	direct object	direct object
personal pronoun	dative object	subject complement
determiner pronoun	subject complement	object complement
non-inflecting pronoun	object complement	prepositional object
finite verb	prepositional object	adverbial object
participle	adverbial object	adverbial
infinitive	adverbial	free nominal adjunct

check it! [New text](#) (back to Linguistics page)

Though the possibility to work with free, real life text and to make up one's own examples is compelling proof of the efficiency of the system's underlying parser (or at least proves that such efficiency is being claimed ...), not all students master a foreign language to such a degree that they enjoy inventing their own sentences, and they will not always come up with

a *correct*. sentence, either. And even copying and pasting from corpus texts (the obvious solution, implemented in both the Portuguese and English modules) may become tedious in the end. On the other hand, many people enjoy a test match - at least as long as they are not being watched or judged. Therefore, I have integrated a sentence randomizer into the system, that offers corpus examples of its own⁸ if the input window is left empty in interactive grammar mode. In order not to hit upon headlines and other unorthodox or "incomplete" text material, all random text choices are cut at sentence delimiters (full stop, colon etc.), and filtered out if they do not contain at least one finite verb.

9. Corpus searches

In the case of persisting difficulties with a particular grammatical topic, or for want of a satisfactory definition, a student may want to look at a few examples of how the concerning feature is used in different sentences, something one would expect to achieve in traditional, text book based, exercises by referring back to a specific chapter in the grammar book. In the case of an IT interface with a live parser at its disposal, there is - in principle - no limit to the amount of corpus text to be searched for "typical" examples, and the illusion of a concise "chapter" can be created even with a chaotic "book" (corpus) with thousands of pages: While the grammar server searches the "book", the terminal will show the "chapter". Let's assume, for instance, that the student has a problem with Portuguese verb chains - he is in doubt just how and if prepositions can be integrated in auxiliary verb structures. He therefore clicks "open corpus search" on the top page, and looks for prepositions preceded by auxiliaries (@FAUX or @IAUX) or followed by post-auxiliaries (@#ICL-AUX<). The system will then output a very long "chapter" on this topic, and he may get the impression, that, say, such verb chains do not occur in infinitive-subclauses. To look for counter-examples, all he has to do is add a form/function label for such subclauses. In (12), the search is for non-finite *subject* subclauses: @#ICL-SUBJ>_PRP_@#ICL-AUX<:

(12) searching for corpus examples

⁸ Presently, the text base for the randomizer is about 1,000 sentences large, but since it is based on automatic analysis, ten or a hundred times that number would not be a problem, either.

```

search pattern: @#ICL-SUBJ> ([^@]*@Æ[^@]*) PRP ([^@]*@Æ[^@]*) @#ICL-AUX<
corpus search: @#ICL-SUBJ>_PRP_@#ICL-AUX<

```

... *para ADVL um >N profissional P< que SUBJ> FS-N< já ADVL> ganhou FMV vários >N prêmios
<ACC por <ADVL as >N campanhas P< de N< marketing P< que ACC> FS-N< desenvolveu FMV , fala
FMV quatro >N idiomas <ACC além <ADVL de A< o >N português P< e CO passou FMV os >N
últimos >N dez >N anos <SUBJ ocupando IMV ICL-<ADVL postos <ACC de N< chefia P< , ##*#> voltar
IAUXICL-SUBJ> a PRT-AUX< folhear IMV ICL-AUX< <##*# a >N seção <ACC de N< classificados P< de
<ADVL os >N jornais P< é FMV uma >N experiência <SC traumática N< ...

... ##*#> *deixar IAUX ICL-SUBJ> de PRT-AUX< responder IMV ICL-AUX< <##*# a <PIV uma >N ligação P<
de N< um >N conhecido P< hoje=em=dia <ADVL pode FAUX significar IMV ICL-AUX< fechar
IMV ICL-<ACC uma >N porta <ACC que SUBJ> FS-N< pode FAUX ser IMV ICL-AUX< útil <SC em A< caso P<
de N< desemprego P< ...

.....

Output conventions:

[WORD CLASS DEFINITIONS](#)
[SYNTACTICAL CATEGORY DEFINITIONS](#)

noun **N**, proper noun **PROP**
personal pronoun **PERS**, "nominal" pronoun **SPEC**, determiner pronoun **DET**
adjective **ADJ**, adnominal participle **PCP**, numeral **NUM**
verb **V**, verbal participle **PCP**
adverb **ADV**, preposition **PRP**, conjunction **KS/KC**, interjection **IN**, affix **EC**

Here, two results of this quite specific search task are given. Note that again, the "enriched text" notation permits text coherence and facilitates context understanding. The particular structure looked for is marked by fat arrows, but thanks to the concise notation, the whole sentence context can be shown together with most of the tags.

Even tags *not* shown, like the flexion category 'plural', the base form 'amigo' or the valency feature 'monotransitive', can be searched for: Virtually any combination of word forms, base forms, inflection tags and syntactic function can be searched for each individual word in any combination of words as well as one or more obligatory or optional dummy words. Obviously, the *real* search pattern for complex searches is much longer than the chain of tags entered by the user. The system automatically "translates" the search parameters into a regular expression string (cp., for instance, the search pattern line at the top of [12]) to be used by fast, specialized search algorithms running on the unix based grammar server.

Let's look at another, English, example, where a student wants to write an essay on a different type of verb chain - causatives, which he believes to exhibit a structure where a causative verb is followed by accusative+to+infinitive. A corresponding search (FVIN_ACC_'to'_INF) will indeed yield a lexicographically interesting list of causative verbs:

(13a) "causative" verbs in English

...But CO he SUBJ added FMV that SUB none SUBJ of N< Georgia's >N congressmen P< specifically ADVL
 ###> asked FMV him OBJ to INFMARK offer IMV <### the >N resolution OBJ....

... Chairman >N C& >N Richard >N Mears SUBJ pointed FMV out ADVL that SUBJ@OBJ perhaps ADVL this
 SUBJ was FMV not ADVL strictly ADVL a >N school >N board >N problem SUBJ@SC , in=case=of ADVL
 atomic >N attack P< , but CO that SUB the >N board SUBJ would FAUX cooperate IMV so=far=as SUB
 possible OBJ to INFMARK get IMV @N< the >N children OBJ to INFMARK where ADVL the >N parents SUBJ
 ###> wanted FMV them OBJ to INFMARK go IMV <###

... Practice SUBJ ###> helps FMV you OBJ to INFMARK get IMV <### your >N timing OBJ@SC @IMV down
 ADVL

... That SUBJ ###> permitted FMV us OBJ to INFMARK start IMV <### controlling IMV the >N ball OBJ
 right=away ADVL "...

... Philadelphia SUBJ ###> permitted FMV him OBJ to INFMARK seek IMV <### a >N better >N
 connection OBJ after SUB he SUBJ had FAUX refused IMV to INFMARK reconsider IMV his >N decision OBJ
 to INFMARK end IMV @N< his >N career OBJ@SC as N< @ADVL a >N player P<

... For ADVL that >N matter P< , Stan >N Musial SUBJ is FMV rare SC , possessing IMV the >N disposition
 OBJ that SUBJ ###> enabled FMV him OBJ to INFMARK put IMV <### out ADVL the >N same OBJ for ADVL
 seven >N managers P< , reserving IMV his >N opinions OBJ , but CO not ADVL his >N effort OBJ@P<

If the student then wants to generalize his structural assumption he may try to admit interfering material between the accusative and the infinitive (VFIN_ACC_?_'to'_INF):

(13b)

...But CO he SUBJ added FMV that SUB none SUBJ of N< Georgia's >N congressmen P<
 specifically ADVL ###> asked FMV him OBJ to INFMARK offer IMV <### the >N resolution OBJ....

... Both >N parties SUBJ@OBJ@P< in N< @ADVL the >N last >N election P< told FMV us OBJ that SUB
 we SUBJ need FMV a >N five >N per=cent >N growth SUBJ@OBJ in N< @ADVL the >N gross >N
 national >N product P< - but CO neither SUBJ@P< ###> told FMV us OBJ how ADVL to INFMARK
 achieve IMV <### it OBJ "...

... I SUBJ ###> find FMV it OBJ hard OC to INFMARK understand IMV @N< <### how ADVL anyone
 SUBJ@OBJ seeking IMV a >N position OBJ in N< @ADVL public >N life P< could FAUX demonstrate
 IMV such >N poor >N judgment OBJ@>N and CO bad >N taste OBJ....

To his possible surprise, he will now encounter quite a few examples of similar non-causative constructions, which he might not have hit upon *without* using the corpus search engine. Here, 'tell' in the second example is not about ordering (i.e. causative) but about giving (information), and the structural difference may be the fact that the infinitive-clause is headed by what I would call a *relative adverbial* ('how'). The third example, finally, is "causative" (or, rather, transobjective), but it is not the infinitive that functions as object complement, but the embedded adjective "hard".

What our student probably *was* looking for are cases like the second example with *non-relative* adverbs. So he may go for a more well defined encore and search for VFIN_ACC_ADV_'to'_INF:

(13c)

```
... Both >N parties SUBJ@OBJ@P< in N<@ADVL the >N last >N election P< told FMV us OBJ that SUB we
SUBJ need FMV a >N five >N per=cent >N growth SUBJ@OBJ in N<@ADVL the >N gross >N national >N
product P< - but CO neither SUBJ@P< ###=> told FMV us OBJ how ADVL to INFMARK achieve IMV <### i
OBJ " ...

... Nobody SUBJ can FAUX deny IMV the >N right OBJ of N< former >N Chief >N Executives P< to
INFMARK take IMV@N< part OBJ in N<@ADVL politics P< , but CO the >N American >N people SUBJ@P<
###=> expect FMV them OBJ always ADVL to INFMARK remember IMV <### the >N obligations OBJ of
national >N leadership P< and CO to INFMARK treat IMV issues OBJ with N< a >N sense P< of N<
responsibility P< ...
```

Finally, the prototypical pre-infinitive adverbial shows up, in the second example: 'always' (another possibility would, of course, be 'never'). However, the matrix verb in question, 'expect', may suggest new subdivisions in what was thought to be a homogeneous group of "causatives" ...

This way, hopefully, student may turn into "researcher", learning grammar by making his own.

10. Conclusion

Experiments with notationally filtered Constraint Grammar analyses for Portuguese and English free, running text have shown that an efficient parser can be turned into a valuable grammar teaching tool, - if it can be accessed through a "non-technical" interface, that honours the four basic principles of "live" teaching: Interactivity, flexibility, naturalness and tutoring. By exploiting the distributed character of the Internet, one or more central grammar servers can service a large number of simultaneously active, individualized versions of the teaching interface, at the same time allowing easy up-dating and solving collateral problems of copy-right, compatibility and accessibility.

As to naturalness, students are allowed to work with free language samples and use the tools they know from other "friendly" software, like windows, mouse and menus. In fact, the interface can be run "single-handedly", by mouse alone, without ever touching the keyboard. As to flexibility, one can choose from different levels of analysis and descriptiveness, and even move between different schools of syntactical description. Users may either ask for a ready analyses or interactively build their own with the computer tutoring their choices, defining terms, translating text and exemplifying concepts. Finally, more research-minded students can venture into the realm of corpus analysis and put grammatical notions to the test.

Paedagogically, I have advocated the advantages of word based form and function markers (tags), flat dependency syntax and in-text meta-notation in the form of colour codes and indexing. Ideally, in the case of "wrong" analyses, students should not be criticized for diverging choices if these are motivated by different grammatical backgrounds. Likewise, - unless the student explicitly asks for it -, testing should not focus on the quantification of errors ("scores"), but on the game aspect of the challenge, i.e. the process as such, not the

result. In this vein, the interface features a sentence randomizer suggesting unknown sentences to the student for interactive analysis.

Finally, integration of the IT teaching tool into the broader context of ordinary, pre-existing language teaching is encouraged. Here, special notational filters on top of Constraint Grammar parsers, as well as text book based closed corpora are possible solutions. This way, given the interface's inherent flexibility, it should not be too hard to introduce similar Internet tools on all levels of language teaching, in universities as well as secondary and primary schools.

References:

- Bache, Carl & Davenport, Mike & Dienhart, John & Larsen, Fritz: *An Introduction to English Sentence Analysis*, Munksgaard 1991, 1993
- Bick, Eckhard: *Portugisisk - Dansk Ordbog*, Mnemo, Århus, 1993, 1995, 1997
- Bick, Eckhard: *Automatic Parsing of Portuguese*, i *Proceedings of the Second Workshop on Computational Processing of Written Portuguese*, Curitiba, 1996
- Bick, Eckhard: *Dependensstrukturer i Constraint Grammar Syntaks for Portugisisk*, in: Tom Brøndsted & Inger Lytje (red.), *Sprog og Multimedier*, Aalborg Universitetsforlag, 1997
- Bick, Eckhard: *Automatisk Analyse af Portugisisk Skriftsprog*, Ph.D. project presentation, forthcoming
- Karlsson, Fred & Voutilainen, Atro & Heikkilä, Juka & Anttila, Arto (eds.), *Constraint Grammar, A Language-Independent System for Parsing Unrestricted Text*, Mouton de Gruyter, Berlin 1995

Lytje, Inger & Donner, Claus: *Gratex. Dynamisk grammatik til dansk sprog*, ORFEUS and Aalborg Universitetsforlag, 1996

Voutilainen, Atro & Heikkilä, Juka & Anttila, Arto: *Constraint Grammar of English, A Performance-Oriented Introduction*, Publication No. 21, Department of General Linguistics, University of Helsinki, 1992