

Grammar for Fun: IT-based Grammar Learning with VISL

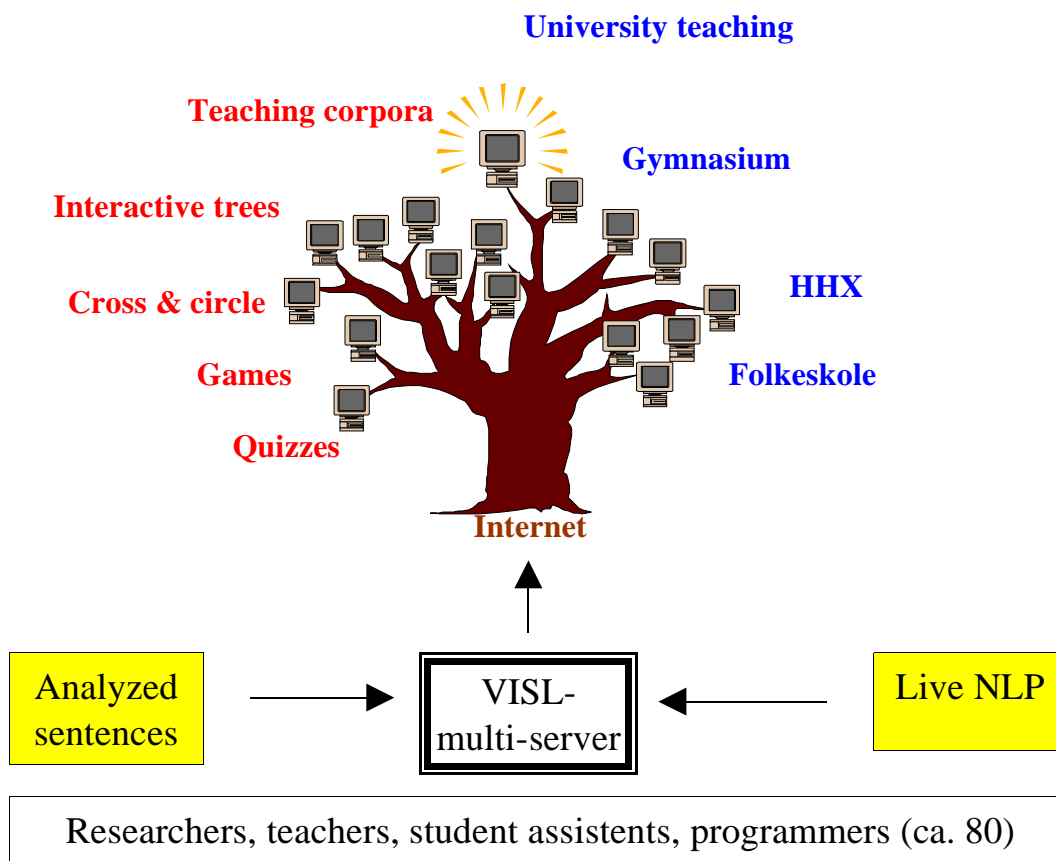
Eckhard Bick

e-mail: eckhard.bick@mail.dk , web: <http://beta.visl.sdu.dk>

1. Abstract

This paper presents an integrated interactive user interface for teaching grammatical analysis on the Internet (Visual Interactive Syntax Learning), developed at the University of Southern Denmark, offering a unified system of analysis for 22 different languages, 7 of which are supported by live grammatical analysis of running text. For reasons of robustness, efficiency and correctness, the system's internal tools are based on the Constraint Grammar formalism (Karlsson 1990), but users are free to choose from a variety of notational filters, supporting different descriptive paradigms, with a current teaching focus on syntactic tree structures, language independent grammatical categories and the form-function dichotomy. VISL's core NLP-programs use hybrid multi-level parsers (Bick, 2003), while teaching applications and corpus searching tools are implemented as platform independent Java-programs and Perl-cgi's. Though lexica and parsing rules are developed individually for each language, a common CG and treebank data format facilitates source data transfer into grammar teaching games, structural or color based visualisation, and linguistic revision of corpus data.

The VISL teaching network



1. Background

When the VISL project started in 1996, its primary goal was to further the integration of IT tools and IT based communication routines into the university language teaching milieu at Odense University (Denmark), and more specifically, to develop tools for Visual Interactive Syntax Learning. The initiative was funded jointly by CTU (Center for Teknologi-Støttet Uddannelse) and Odense University for 3 years, and has since grown into a multi-language conglomerate of independently funded subprojects. Early on, a decision was made to develop both a "closed" and an "open" system, and to tailor teaching applications for maximal synergy, such that they would be able to take input from both the closed and open language data sources, - and do so in a largely language independent way. Using a common descriptive format, hand annotated treebanks were built, covering between 100 and several thousand pedagogically chosen sentences, depending on the target language. Larger, automatically created and revised research treebanks have since been added for Danish, Portuguese and French, and for 7 languages, new teaching treebanks can be created semi-automatically, allowing schools to enter their own materials into the VISL system.

2. A unified approach to grammar

The central principle of VISL's language analysis is its focus on surface structure (expressed as either dependency relations or syntactic tree structures) and the form-function dichotomy. Following Bache et.al. (1993, 1999), function symbols start with upper case letters, form symbols with lower case letters, and both are combined in a combined colon-separated "symbol" (text) or a function-over-form graphical tree node. For the dependency notation, international CG conventions are followed, with upper case letters for all primary tags, using the @-symbol to introduce function tags, and arrow heads (>,<) or ID-numbers for head oriented dependency markers.

VISL lite vertical tree (non-graphical notation, filtered)	VISL vertical tree (non-graphical notation, incl. morphology)
UTT:cl	STA:fcl
S:prop VISL	S:prop("VISL") VISL
P:v er	P:v-fin("være",pr,akt) er
Cs:g	Cs:np
=D:art et	=DN:art("en",neu,sg,idf) et
=H:n forskningsprojekt	=H:n("forskningsprojekt",neu,sg,idf,nom) forskningsprojekt
=D:cl	=DN:fcl
==S:pron der	==S:pron-rel("der",nG,nN,nom) der
==P:v involverer	==P:v-fin("involvere",pr,akt) involverer
==Od:g	==Od:np
===D:pron mange	===DN:pron-indef("mange",nG,pl,nom) mange
===D:adj forskellige	===DN:adj("forskellig",nG,pl,nD,nom) forskellige
===H:n sprog	===H:n("sprog",neu,pl,idf,nom) sprog

VISL	[VISL]	<heur> <*>	PROP NOM	@SUBJ> #1->2
er	[være]	<vk>	V PR AKT	@FMV #2->0
et	[en]		ART NEU S IDF	@>N #3->4
forskningsprojekt	[forskningsprojekt]		N NEU S IDF NOM	@<SC #4->2
,				#5->0
der	[der]	<rel>	INDP nG nN NOM	@SUBJ> #6->7
involverer	[involvere]	<vt>	V PR AKT &MV	@FS-N<
#7->4				
mange	[mange]	<quant>	DET nG P NOM	@>N #8->10
forskellige	[forskellig]		ADJ nG P nD NOM	@>N #9->10
sprog	[sprog]		N NEU P IDF NOM	@<ACC #10->7
.				#11->0

Meeting regularly over 4 years, the VISL group of university teachers has invested considerable effort in discussing the compatibilities, incompatibilities and blind spots of different national and linguistic grammar traditions, and agreed upon a common superset of symbols, as well as conventions for adding subcategories - the so-called VISL cafeteria. In 2001, a reduced symbol set for propedeutic use and schools, "VISL lite", was introduced, and the Danish cross-and-circle system adapted to match the function categories used in VISL. Currently, 4 different levels of complexity are distinguished, all automatically derived from a common super-format. At the *lite* level, 11 word classes and 14 primary functions are used. *Ultralite* fuses, in addition, most non-inflecting word classes into a "nil"-category.

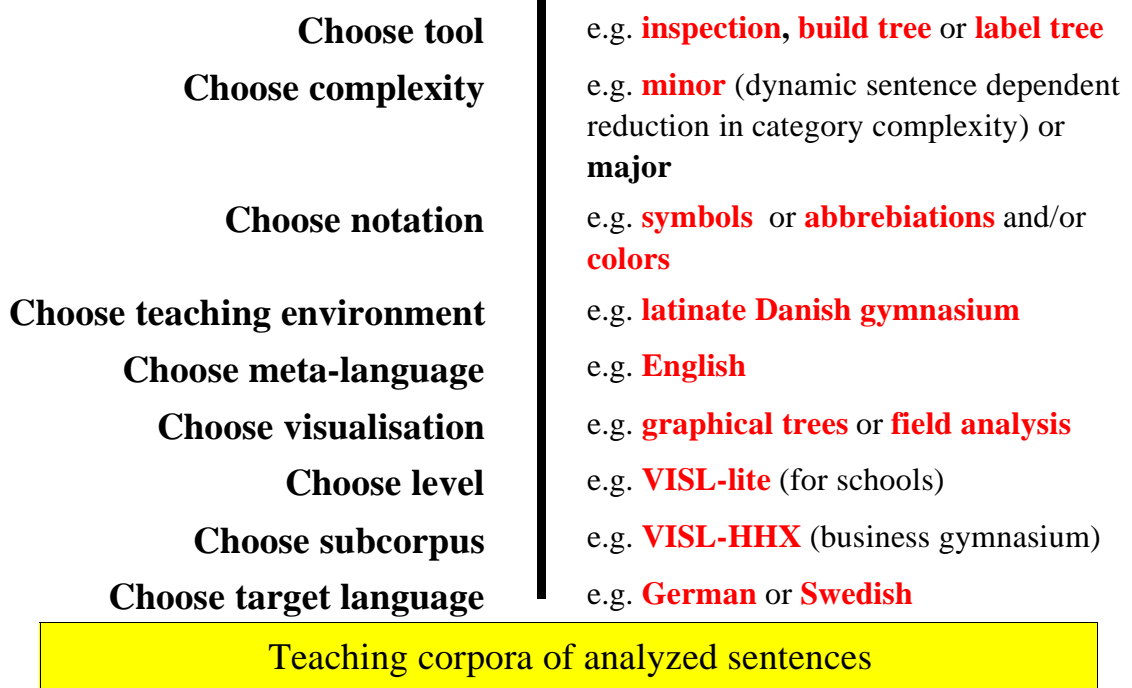
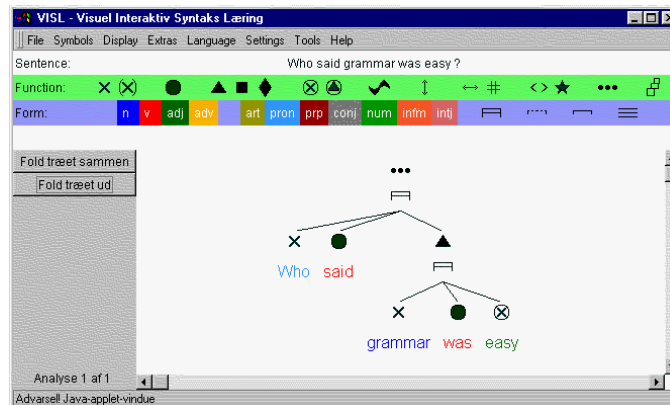
- Predicator (P), Verbal (V)
 - Auxiliary (Vaux), also as (D)
 - Main verb (V*, Vm), also as (H, K)
 - Verb chain particle (Vpart), also as <> (D), simplified as (A)
 - Infinitive marker (Vi, INFM), also as <> (D)
- ✕ Subject (S)
 - Formal or provisional subject (Sf), possibly with the subclass of situative subject (Ss)
- ▲? Direct (accusative) object (Od)
 - Formal or provisional object (Of)
- Indirect (dative) object (Oi)
- ◆ Prepositional object (Op), at the lite level filtered into (A)
- ✕ Subject complement (Cs), Subject predicative (Ps)
 - Free subject predicative (fPs, fCs), simplified as (A)
- ▲ Object complement (Co), Object predicative (Po)
 - Free object predicative (fPo, fCo), simplified as (A)
- ~ Adverbial (A), with possible subdivision of free (fA) or bound (bA, bAs, bAo)
- * Head (H), Kernel (K)
- <> Dependents (D)
 - Subordinator (SUB)
- ↔ Co-ordinator (CO)

Conjunct (CJT)

3. Internet based teaching tools

Though on the decrease, and not necessarily shared by their students, there is still some resistance among both school and university language teachers with regard to all things technical. There is, of course, a fundamental difference in terms of "technicality" 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 computer program can "know" more - in terms of facts and examples, and within a well-defined subject matter - than a human teacher. And last, but not least, a computer system, especially if accessible through the internet, can teach an unlimited number of students at the same time in what optimally still amounts to an individual manner.

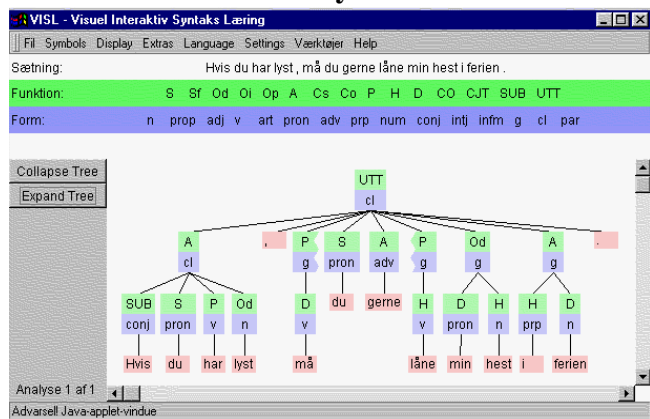
3.1 Flexibility



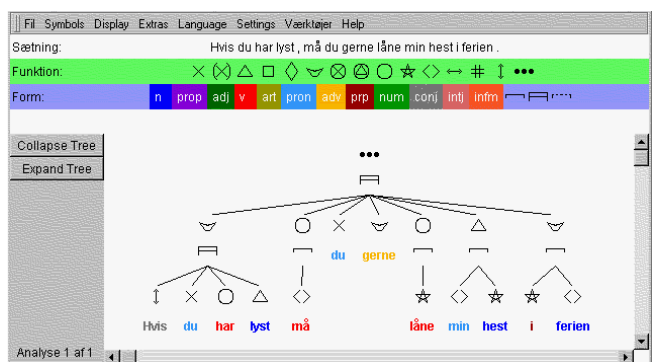
The VISL interface is notationally flexible, i.e. the user can choose between several notational conventions (e.g. color or symbolic encoding, word based tag mode, enriched text, tree structures), switch to another meta-language, and move back and forth between different levels of complexity. For instance, depending on the exercise chosen, the type and number of grammatical categories used may be changed, or even set to depend on the individual sentence ("minor" mode). In order to make work more colourful, it is also possible to move between text book material, copied "live" texts, randomized test sentences and one's own creative idiolect.

In the tree structure example below, the user can switch back and forth between letter symbols and graphical symbols, more than double the number of categories, or reduce the tree to a pure function tree (green only).

Letter symbols



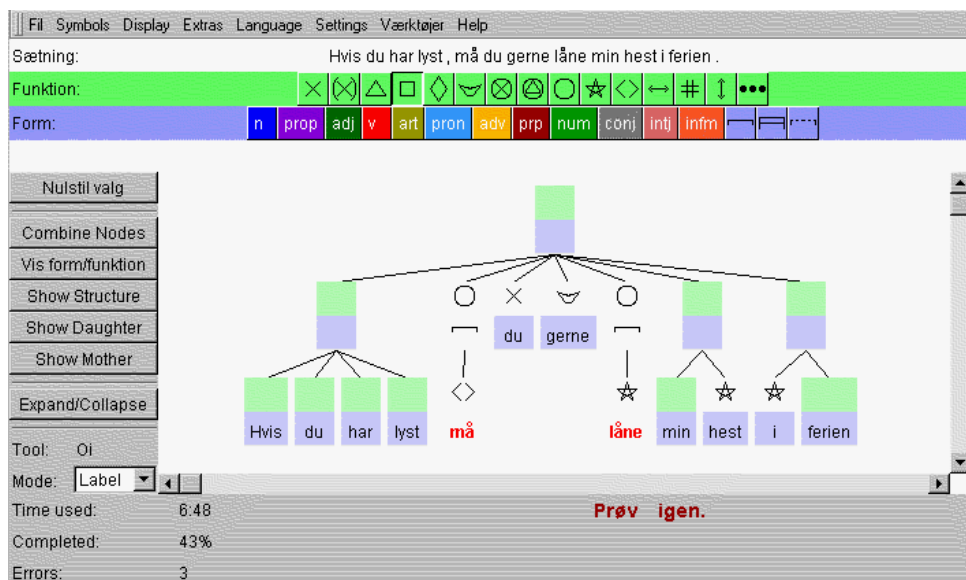
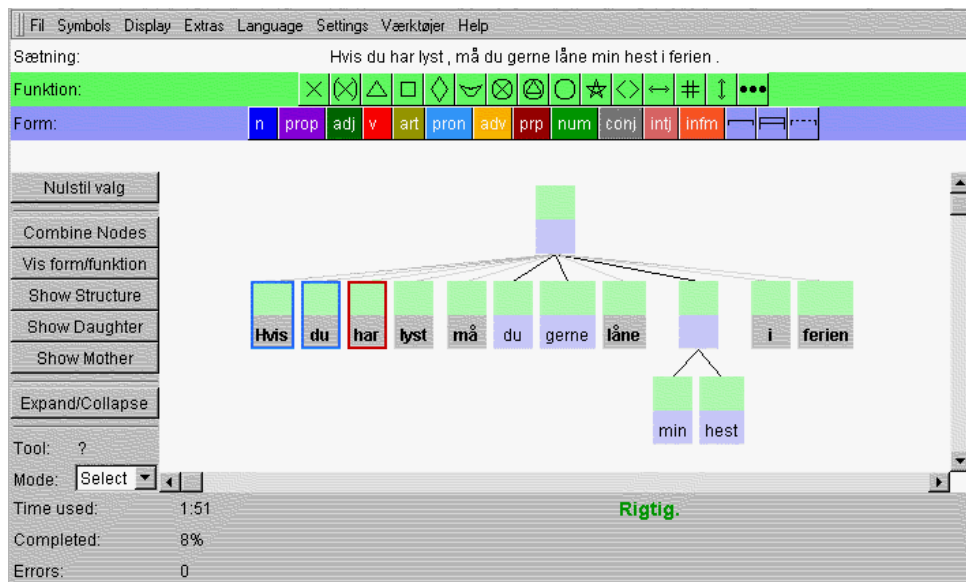
Graphical symbols



VISL's unique integration of teaching and research tools would even allow the user to experiment with different kinds of subjects or add a couple of place and time adverbials and rerun the sentence in free-text mode – with exactly the same graphical setup and paedagogical functionality.

3.2. Interactivity

VISL's java-tree interface for grammatical analysis allows the step-by-step interactive inspection, construction and labelling of syntactic trees using menus, mouse clicks and drag-and-drop movements, all known from basic text processor functionality. In the first example, a student has recognized the np "min hest", but has yet to assemble "lyst" onto the predicator ("har") of the adverbial subclause to the left.



When a sentence proves problematic or incomprehensible, the user can modify it, or ask for the computer's opinion (show-me option). In grammar games like *Paintbox*, *Post Office* or *Shoot-the-Verb*, interactivity integrates a certain element of competition, and is further enhanced by sound effects, timers and high-scores.

3.3. Naturalness

A major draw back of most language teaching software (or, for that matter, language analysis software) is that they do not run on free, natural language, but only on a small set of predefined sentences or structures ("toy lexica" or "toy grammars"), that cannot be modified or replaced. In the VISL interface, for better or worse, the underlying lexica and grammars cover the whole language, supporting gradual and comparative changes in a given sentence, or confronting the user with the stimulating lexical freshness and structural unpredictability of running natural text.

The second aspect of naturalness concerns, as mentioned above, "untechnical" ergonomics, and as much keyboard-interaction as possible has therefore been

replaced by graphical and mouse governed tools, like menu choices 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.

3.4. Tutoring

Tutoring is traditionally a human task, and difficult to simulate in a computer interface. Therefore, in its strict sense, it has been one of the last features to be implemented on the VISL site. A certain minimum of tutoring can be achieved simply by providing guided tours, help windows, clickable definitions of grammatical terms, show-me-buttons, and ready access to topic conditioned corpus examples (through VISL's corpus search site). However, real tutoring asks for more specific and individual comments. Therefore, to help students with the tree-building and -labelling task, we have begun to implement so-called error-comment files, where pedagogical remarks (and suggested reading-links) are stored for all common and some rarer combinations of "correct label expected" and "wrong label chosen", as well as for different types of wrong attachment (phrase and clause grouping).


4. Grammar Games

VISL employs grammar games as part of a complexity based teaching progression. Word class games are used to support teacher based explanations and, for instance, blackboard or paper based underlining or slot filler exercises. *Paintbox* is a quiet word class coloring game which can be played with any number of word classes, while *Labyrinth* and the Tetris-based *WordFall* are faster and more competition oriented, using a penalty system and allowing high-scores. The latter is an example of a game that will evaluate user performance and offer tailor made re-runs.

<i>Paintbox</i>	<i>WordFall</i>	<i>Labyrinth</i>
		

At the syntactic level, a fundamental decision is made between *word based* and *constituent based* use of function categories. The latter, probably closer to a child's intuitive, semantics based understanding, will assign subject function to the *head* of a syntagma, i.e. "reduce" *the little nice rabbit, who ate the carrots* to its core, "rabbit". The result is a flatter, more immediate syntax, well-suited to the Danish cross-and-circle system, and implemented in the *PostOffice* game, where words are stamped for function. In VISL's teaching progression, this game will be used *before* constituent based games like the "action" game *SpaceInvaders*. A compromise is

SynTris, where there is a function-only mode, with constituents shown but not interactively built.

<i>PostOffice</i>	<i>SynTris</i>	<i>SpaceInvaders</i>
		

A new development are morphology games, such as *BalloonRide* and the 2-player *TrainRace*, the usefulness and difficulty of which will, of course, depend upon the number and size of morphological paradigms in a given language.

5. Evaluation

Though numerous teaching institutions, as well as the Ministry of Education through the years have accepted the worth of VISL at face value, or rather, concluded from its IT-promoting and interactivity ("fun") values to its objective educational usefulness, most research has gone into developing and evaluating NLP and applicative tools for VISL, rather than evaluating their impact on students' grammatical knowledge, and only a few pilot projects have addressed the latter area, such as the Syntax Course evaluation for English at Odense University, and the Norwegian 7th and 8th grade study, done when establishing the GREI front-end for Norwegian VISL users (<http://www.tekstlab.uio.no/grei/>).

Another aspect of evaluation is of interest for individual teacher-users: Is a given student progressing in a given area of grammatical training or not? Can class-wide tendencies or problems be observed?

During the first half of 2004, VISL has created a first version of a tool that will hopefully address both evaluation aspects at the same time. The tool itself, christened *KillerFiller*, is basically an IT-based pandemonium of slot-filler exercises, drawing random sentences not only from VISL's teaching corpora, but also from its research data bases, and removing words (e.g. all prepositions or verbs), which the user has to fill back in. Depending on the language, for some word classes grammatical categories or base forms will be provided to avoid ambiguity. In order to handle evaluation, users are assigned login ID's and passwords, and a server-side database stores the results from every run, sorted by language, user and exercise type. After each run, improvement statistics and graphs are shown, and teachers and evaluators can access historical overview pages for relevant sections of the stored data. Thus, the tool not only allows to grade the individual user, but also to quantify, say, the average improvement after a VISL based grammar course.

<p>Please login to your VISL-game account If you do not have an account, create a new one by clicking here!</p> <p>Username: <input type="text" value="learner"/></p> <p>Password: <input type="password"/></p> <p><input type="button" value="Login"/></p>	<p>Which language do you want to train?</p> <p></p> <p></p>	<p>Sentence collection: <input type="text" value="Grammy 1"/></p> <p>Word class: <input type="text" value="v"/></p> <p><input type="button" value="Show sentence"/></p> <p></p>
---	---	---

□

Kasparow zu (besiegen) (müssen -pr-) für den
Computer ein Genuß (sein) (sein)

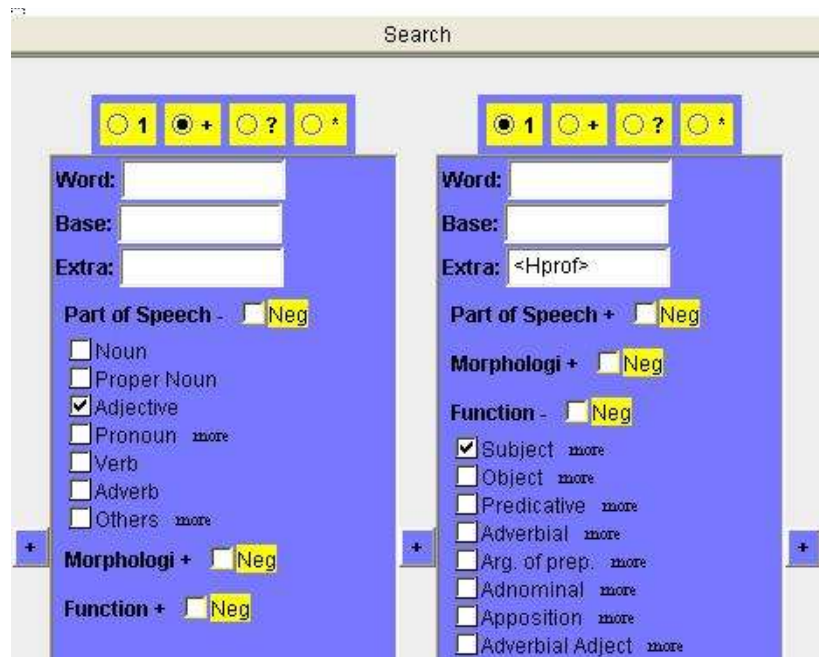
6. Corpora

Transcending its original target area (internet based grammar teaching tools), VISL has generated a number of collateral spin-off results both technological and linguistic. Thus, a number of comprehensive bilingual lexica, valency-lexica and semantic prototype lexica are under development for several languages, and GNU-licence compilers for CG and PSG are being made available to the public. Serving both research and teaching, VISL's corpus site offers a user-friendly search interface, where text corpora are accessible in both raw and annotated form for VISL's core languages. Separate sub-projects are or have been the annotation of the 50 million word Danish Korpus90/2000 (in cooperation with DSL, Denmark), the Danish multi-format Arboretum treebank (200.000 words revised), multilingual Europarl annotation, the French Freebank and a 2 million word tree bank for Portuguese (in cooperation with the AC/DC-project, Oslo). In all, annotated corpora are available for 7 languages.

□



Drawing on the principle of servicing "non-technical" users, corpus data is accessible through a largely menu-based interface not only to researchers, but also to teachers with little or no knowledge of annotation schemes, regular expressions and the like.



From a teaching perspective, corpora allow easy excerption of real life examples for a given syntactic structure, and support do-it-yourself exercises targeting general language awareness, e.g.

- find fixed expressions involving animal words!
- find singular nouns without articles or adjectives (mass nouns?)
- find relative clauses ('der' or 'som'?)

Since the interface also provides alphabetical ordering and statistical evaluation of search result concordances, it is possible, for instance, to quantify the use of loan-words, observe language change, or compare gender-dependent usage.

Finally, it is possible to integrate "live corpora" with teaching applications. Thus, the *TextPainter* offers live analysis of cut-and-paste text in 7 languages. Results can be highlighted for a given category or category combination, say *objects*, *subjects*, *verbs* or *adjectives*. Thus, a sample text can be used to grade a novel as a verb-heavy action text or as an adjective heavy descriptive text. In interactive mode, users have to *find*, say, all objects themselves, having their performance evaluated in terms of an integrated recall/precision measure, the F-score.



7. Spreading and adapting VISL to different user groups

Though originally a university project, VISL has had nationally and Nordic funded co-operative projects for most Danish teaching levels, and has committed itself to maintain servers and software even after the official end date of a project.

- VISL-Gym (2-3 years), with language representatives from the Ministry of Education, involving all 10 languages taught in Danish high schools
- VISL-HHX (2 years), targeting English and Danish in Business high schools
- VISL-SEM (ongoing), targeting teacher training seminaries
- VISL-Folke (ongoing), targeting Danish primary schools
- URKAS, aiming at creating dedicated teaching material for a new high school subject, "Almen Sprogforståelse"
- GREI, Norwegian VISL front end
- PaNoLa (2-3 years), creating teaching treebanks for the major Nordic languages
- PaNoLa-plus (ongoing), creating teaching treebanks for the small Nordic languages

While one aspect of such product simply is the creation of linguistic material, data annotation, harmonisation and filtering, some projects have focused on the didactical integration of existing material. Thus, the grammar course *Grammy* (Bick 2001) suggests a pedagogical progression and text-book support for VISL's online tools.

Bibliography:

- Bache, Carl et. al. (1999). *English Sentence Analysis*, København: Gyldendal
- Bick, Eckhard (1997) *Internet Based Grammar Teaching*, in: Christoffersen, Ellen & Music, Bradley (eds.), *Datalogvistisk Forenings Årsmøde 1997 i Kolding, Proceedings*, pp. 86-106. Kolding: Institut for Erhvervsprog og Sproglig Informatik, Handelshøjskole Syd
- Bick, Eckhard (2001), *Grammy i Klostermølleskoven - "VISL light": Tværproglig sætningsanalyse for begyndere* (Teaching Manual), <http://visl.sdu.dk/visl/light>
- Bick, Eckhard (2003), [A CG & PSG Hybrid Approach to Automatic Corpus Annotation](#), In: Kiril Simow & Petya Osenova (eds.), *Proceedings of SProLaC2003* (at Corpus Linguistics 2003, Lancaster), pp. 1-12
- Dienhart, John (2000), *VISL-projektet: Om anvendelse af IT i sprogundervisning og -forskning*. In: *At undervise med IKT*, pp. 51-70. Gylling: Narayana Press
- Karlsson, Fred (1990). *Constraint Grammar as a Framework for Parsing Running Text*. In Karlgren, Hans (ed.), *COLING-90: Papers presented to the 13th International Conference on Computational Linguistics*, Vol. 3, pp. 168-173. Helsinki: RUCL