A Web-based Instructional Platform for Constraint-Based Grammar Formalisms and Parsing

Detmar Meurers
Ohio State University
dm@ling.osu.edu

Gerald Penn
University of Toronto
gpenn@cs.toronto.edu

Frank Richter
Universität Tübingen
fr@sfs.uni-tuebingen.de

ACL Workshop on
Effective Tools And Methodologies for Teaching NLP and CL
Philadelphia. July 7, 2002

Overview

- Why are we interested in web-based teaching?
- Feature structures as a unifying theme
- Seminar-style courses as content basis
- From seminar-style courses to a web-based teaching framework
  - Problems with seminar-style courses
  - What's involved in our approach
  - Education and collaborative learning technology research
  - Developing the framework
    - Three core modules
    - Use of hyperlinks and visualization
    - Outlook

Why web-based teaching?

- Computational linguistics is an interdisciplinary field incorporating knowledge of:
  - linguistics
  - discrete mathematics and statistics
  - algorithms and data structures
  - application domains

- The CL student audience comes from a wide variety of backgrounds (languages, linguistics, computer science, engineering, etc.)

$\Rightarrow$ "Buffet-style" learning allows for individualized content and pace

Building a teaching framework around a unifying theme

- Feature structures are an attractive unifying theme since they
  - underly the most comprehensive and rigorous syntactic theorizing in linguistics (HPSG, LFG, TAG, . . .)
  - are used in most grammar implementation efforts and thereby allow discussion of algorithms involved in processing such grammars
  - have clear formal foundations

$\Rightarrow$ Joint project on *Grammar Formalisms and Parsing* funded by the German Federal Ministry for Education and Research (BMBF) as part of the consortium *Media-intensive teaching modules in the computational linguistics curriculum (MiLCA)*.

Seminar-style courses as content basis

- Constraint-based grammar implementation in ALE and ConTroll
- Introduction to theory-driven CL
- Model-theoretic introduction to syntax

Constraint-based grammar implementation

Edinburgh ALE course (1993–1997, C. Matheson),
various courses based on ConTroll (e.g., ESSLLI 97, LSA 99, U. Tübingen):

- Hands-on experience for linguists interested in the formalization of linguistic knowledge in constraint-based grammar formalisms.
- Taught in an interactive fashion in a computer lab, combining background lectures with practical exercises + individualized grammar projects at end.
- Background lectures introduce the relevant mathematical, computational, and linguistic background.

Introduction to theory-driven CL

Yearly introduction to symbolic CL for graduate and adv. undergraduate students at OSU, complemented by an introduction to data-driven CL.

Covers the basic issues of the following topics: finite state automata and transducers, formal language theory, computability and complexity, recognizers/parsers for context free grammars, wfs tables, active charts, parsing with complex categories (term unification, graph unification).

The theoretical material is combined with

- practical exercises in Prolog implementing the discussed material (encoding/traversing automata, parsing algorithms, etc.), and
- final individual projects consisting of building and testing grammars for a short English text of the student's choice.

Model-theoretic introduction to Syntax

First introduction to HPSG syntax in Tübingen, for students without prior knowledge in syntax, HPSG or logic.

Course focuses on the basic notions of syntax: parts of speech, phrase structure, agreement, raising, control, unbounded dependencies, binding.

Course starts by introducing a standardized version of the logical description language of HPSG, accompanied with problem sets that require the students to construct three dimensional feature structure models of toy grammars (using feature structures made of Styrofoam and wires).

From seminar-style courses to web-based teaching
Problems with the mentioned seminar-style courses

- Seminar-style teaching format presupposes fairly coherent audience.
- Student/teacher ratio not scalable — computers only used as a medium to implement grammars.
- While theoretical material (overheads and research papers) is in electronic form, content not easily accessible without lecture.
- Lectures make little use of graphical and interactive visualization of formal topics (e.g., compared to *Turing's World*).
- Course follows single path through material — difficult for student to adapt according to specific interests/background.

From seminar-style courses to web-based teaching
Our approach

- Idea: Extend from seminar-style to web-based, self-paced format.
- But, preserve positive aspects of the successful courses:
  - hands-on grammar implementation system
  - integrate syntax, grammar implementation, algorithms, and formal foundations.
- Improvement of these aspects:
  - TRALE system: transparent parsing system with close connections to HPSG linguistics — documented source code integrated with course material to demonstrate algorithmic aspects
  - Large English grammar as illustration of linguistic and computational feasibility ($\Rightarrow$ refinement of CSLI's English Resource Grammar)
- Additional essential prerequisite: Scientific understanding of the nature, possibilities, and limits of web-based learning.

Education and collaborative learning technology research (Lin et al., 1995; Nonaka, 1994; Scardamalia and Bereiter, 1993)

- Balance — distinguish public/private knowledge
- Contribution and notification — extensive linking/annotation; insulated modules prevent debate and revision
- Source referencing — preserve boundaries of idea, indicate credit and historical antecedents
- Storage and retrieval — present contributions in context of those with significant overlap
- Multiple points of entry — where possible, project formal content into transparent, intuitive domain; for differing backgrounds
- Coherence-producing mechanisms — feedback on modules that are "fading"
- Links to external resources — situate contributions in wider context (ILIAS).

Developing the framework
(In progress)

Three core modules: Formal foundations, Constraint-based grammar implementation, Parsing.

Involves:

1. Course content design,
2. Interleaving and hyperlinking of textual materials,
3. Development of graphical, interactive solutions for presenting and interacting with the content of the material,
4. Instructional and reference grammar development, and
5. TRALE system development.

Textual material includes: instructional notes, reference guides, system documentation, annotated system source code, and annotated grammar source code.

Use of hyperlinks

Within the course material, give first-class status to transfer of knowledge among linguistic, computational and mathematical sources.

We distinguish the following kind of links:

- Conceptual/taxonomical — connect instances of key concepts with their definitions
- Empirical context — connect instances of mathematical structures, design decisions and algorithms to encyclopaedic discussions of their motivation/significance
- Denotational — connect linguistic constructions and correctness criteria of algorithms to mathematical definitions that formalise them
- Operational — connect mathematical definitions and their linguistic discussions to instructional material on related algorithms
- Implementational — connect algorithms to annotated TRALE source code.

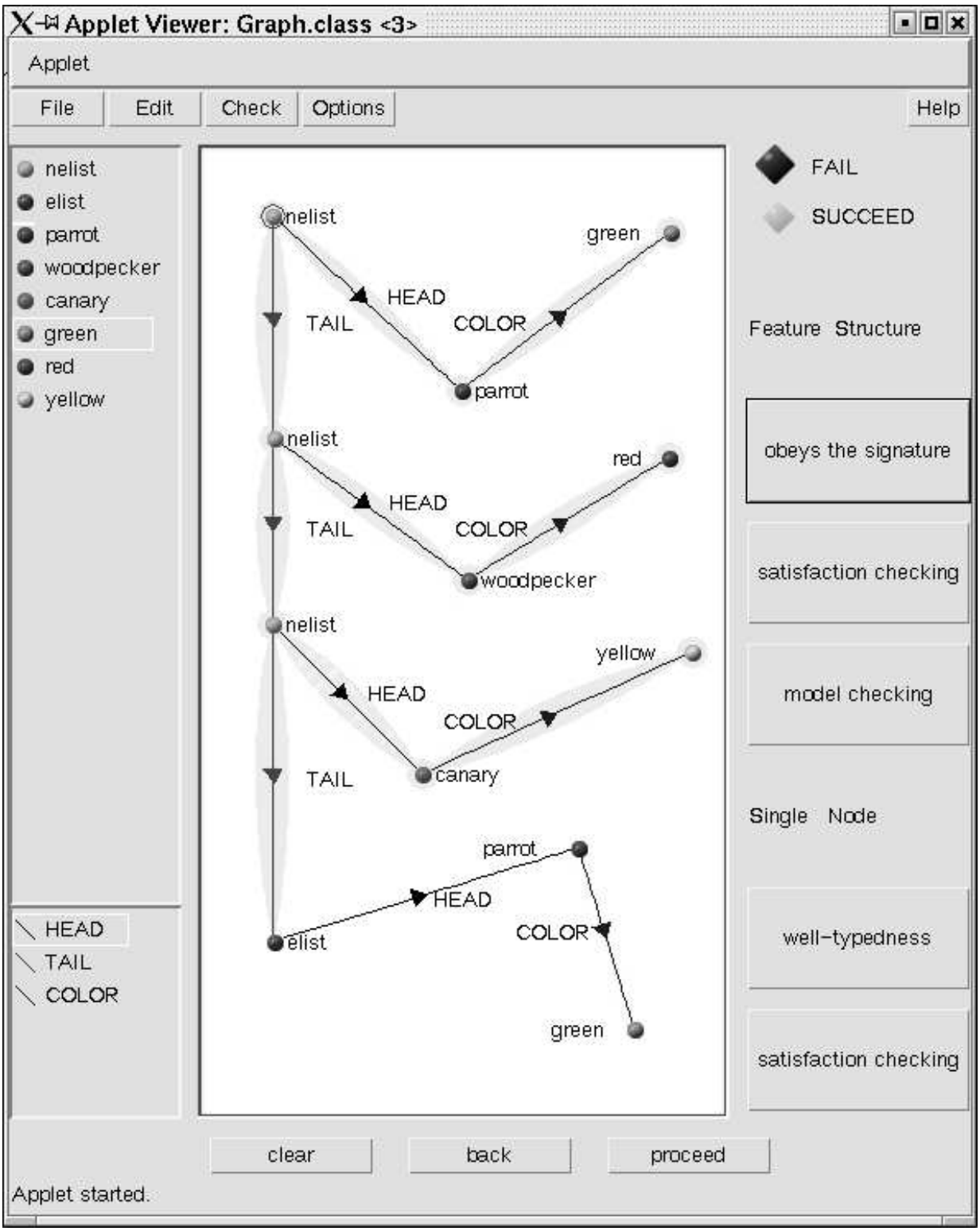
Visualization

The three core modules make use of a number of graphical user interfaces:

- interleaved visualization and interaction with trees and attribute value matrices
- presentation of lexical rules and their interaction
- Emacs-based source-level debugger
- graphical exploration of the formal foundations of typed feature logic (MorphMoulder, MoMo)

The goal of MoMo is to project the formality of its subject, the formal foundations of constraint resolution over typed feature structures, onto a graphical level at which it can be grasped more intuitively.

Graphically evaluating well-typedness



Summary and outlook

- Web-based instructional platform addresses many concerns of teaching interdisciplinary subject like CL to students with diverse backgrounds
- Feature structures can serve as unifying theme to present formalised and implementable linguistic work
- Currently designing:
  - core module on formal foundations
  - core module on constraint-based grammar implementation
  - core module on parsing
  - TRALE system
  - reference grammars, including cleaner ERG
- Available at end of 2003
- Can also serve as open-source standard for future HPSG work.

References

Lin, X., J.D. Bransford and C.E. Hmelo (1995). Instructional Design and Development of Learning Communities: an invitation to dialogue. *Educational Technology* 35(5), 53–63.

Nonaka, I. (1994). A Dynamic Theory of Organizational Knowledge Creation. *Organizational Science* 5(1).

Scardamalia, M. and C. Bereiter (1993). Technologies for Knowledge-building Discourse. *Communications of the ACM* 36(5), 37–41.