

# Functional linguistics and human language technology: new opportunities — or has SFL missed the boat?

John Bateman

ISFC30 Lucknow, Friday 12th December 2003



Universität Bremen

# Overview of talk

- Interactions between SFL and computational approaches to language
- The emergence of “human language technology” as a development area
- After corpora: parsed corpora and multi-treebanks
- Tools for the working linguist: open architectures and interoperability



# SFL and computation: interactions

Halliday (1956) “The linguistic basis of a mechanical thesaurus...”

Halliday (1962) “Linguistics and machine translation”

Henrici (1966) “Notes on the systemic generation of a paradigm of the English clause”

Fawcett (1973) “Generating a sentence in systemic functional grammar”

Davey (1974) “Discourse production: a computer model of some aspects of a speaker”

McCord (1977) “Procedural systemic grammars”



# SFL and computation: interactions

Mann/Matthiessen/Halliday (1983) The Penman text generation system and the Nigel grammar

Cummings (1985) “A PROLOG parser-generator for systemic analysis of Old English nominal groups”

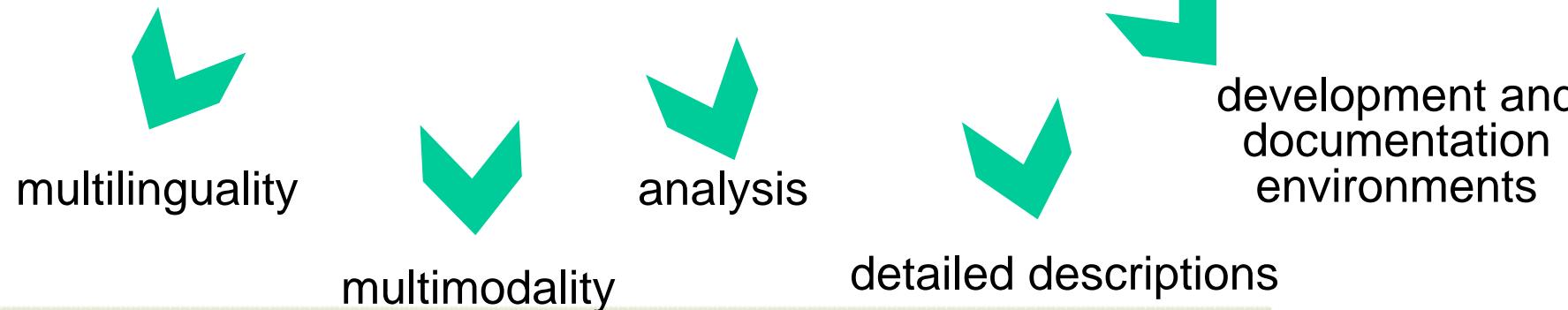
Fawcett/Tucker (1988) Communal text generation system and the Cardiff Grammar



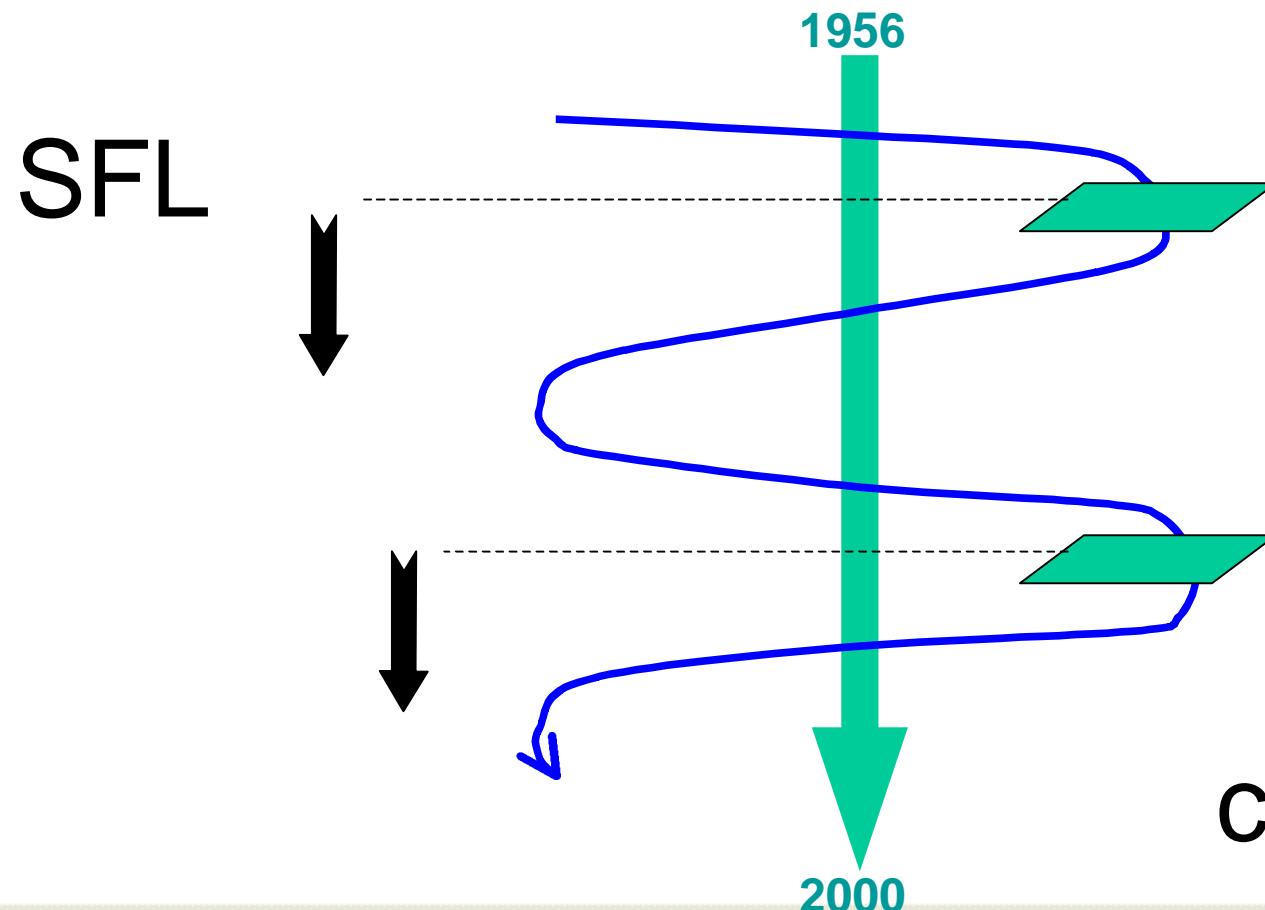
# SFL and computation: interactions

Marilyn Cross (1992) 'Horace'  
Mick O'Donnell (1994) 'Wag'  
Chris Nesbitt (1994) 'HyperGrammar'  
Petie Sefton (1995) interaction  
Licheng Zeng (1993-96) 'Multex'

Elke Teich (1999) 'Komet'  
Liesbeth Degand (1996) Dutch  
Brigitte Grote (1996-) German  
Gordon Tucker (1995) adjectives



# Interactions

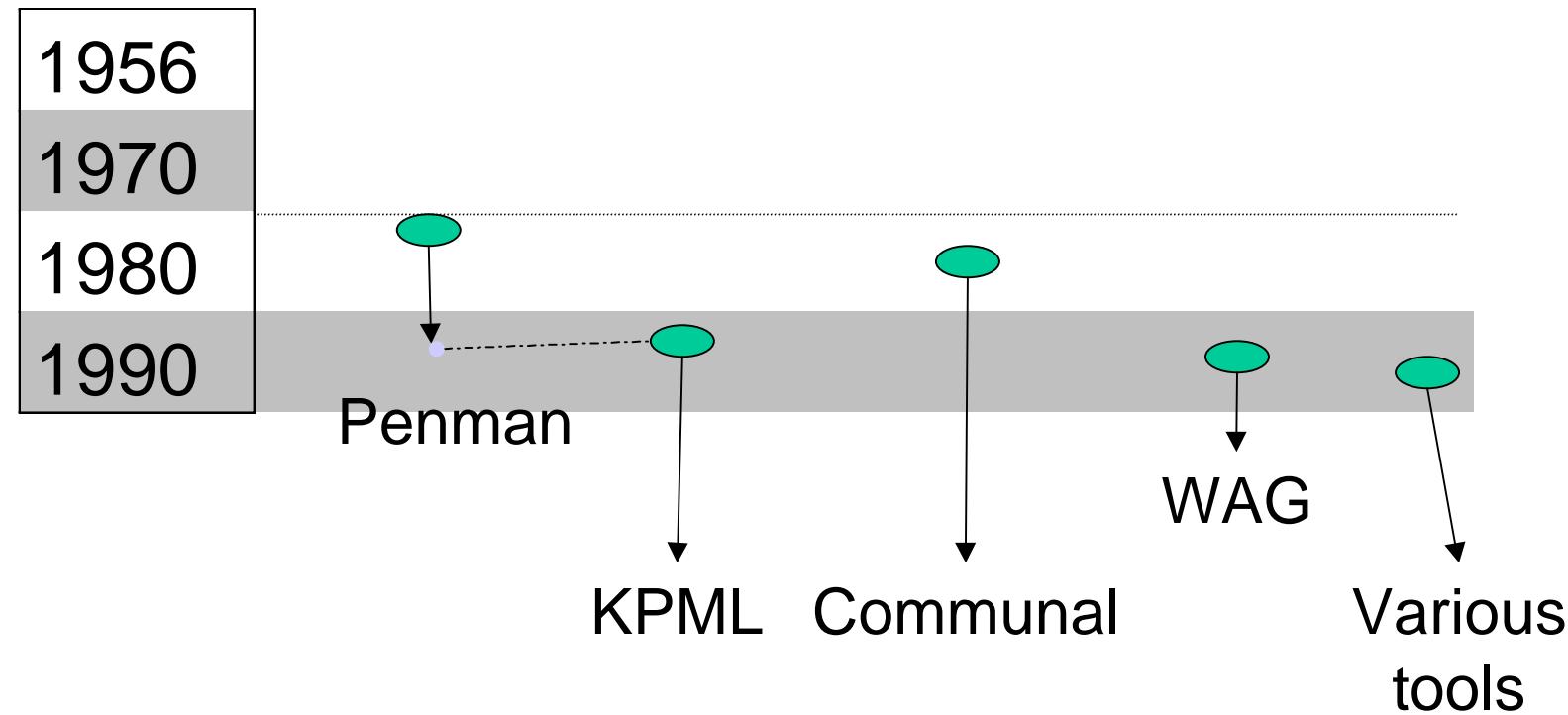


# The state of computation at each point of interaction

1956	theoretical notion of an algorithm, virtually no memory, extremely limited interaction with user
1970	limited memory, slow, basic user interaction
1980	memory available (but expensive), slow, beginnings of varied user interaction: graphical interfaces and debugging environments
1990	more memory available (price dropping), speed increasing, varied possibilities for interaction



# Computational SFL systems



# The state of computation at each point of interaction: 2000

2000      **memory:** for all practical purposes, limitless  
**speed:** for natural language processing tasks, fast  
**interaction styles:** graphical, textual, touch, ...

- Human Language Technology
- Language Engineering
- Linguistic Engineering



# Human Language Technology

- industrial interest in language applications
- substantially larger budgets
- many research and development groups in both universities and companies
  - large lexicons
  - large thesauri (e.g., EDR, Wordnet)
  - ever larger corpora of different kinds of language



# Human Language Technology

- industrial interest in language applications
- substantially larger budgets
- many research and development groups in both universities and companies
  - large lexicons
  - large thesauri (e.g., EDR, Wordnet)
  - ever larger corpora of different kinds of language



# Human Language Technology: Impacts

- There are a considerable number of applications and tasks that can be addressed by a combination of:
  - relatively simple techniques
  - very large scale source data
- Examples
  - speech synthesis
  - information retrieval (also multilingual)



# Human Language Technology: emphases

- much more interest (need) for large-scale handling of material **automatically**
- very substantial efforts (EU, India, ...) on collecting **multilingual** language data
- great concern with **evaluation** and **evaluation criteria**
- notions of ‘best practice’ and **standardization** (both actual and de facto)



# Information vs. Meaning

- Scale alone does not create meaning
- Result:
  - particular ways of structuring information in order to make aspects of its ‘meaning’ more accessible
  - particular techniques for processing such ‘structured data’



# An example of ‘adding meaning’: corpora and annotations

- Just placing a few hundred million words in a computer file does not mean that one has a useful research resource
- Necessary to support the search for significant patterns
- **Development:** combination of corpora and ‘mark-up’ or annotation technology



# An example of ‘adding meaning’: corpora and annotations

- Just placing a few hundred million words in a computer file does not mean that one has a useful research resource
- Necessary to support the search for significant patterns
- **Development:** combination of corpora and ‘mark-up’ or annotation technology



# Three steps...

- Corpora: from raw text to marked-up text
- Text encoding in general
- Corpora: from marked-up text to structured data



# Step 1

- Corpora: from raw text to marked-up text
- Text encoding in general
- Corpora: from marked-up text to structured data



# The problems of searching...

Question: search for bad weather in the novel X...

- (1) select some useful words: *storm, rain, gale, wind*
- (2) search and count
- (3) Results: *storm (32), rain (108), gale (75), wind (345)*

strained, restraint, drain

windlass, windward, tradewinds

But: if we were looking for ‘to **wind**’ in a different sense, we would not then find “**wound**”.



# Example of tagged text from the BNC

Garside, R.G., Leech, G.N., and Sampson, G.R. (eds) (1987). *The Computational Analysis of English: A Corpus-based Approach*. Longman, London.

```
<s c="0000002 002" n=00001>When&AVQ-CJS;  
Captain&NP0; Pugwash&NP0; retires&VVZ;  
from&PRP; active&AJ0; piracy&NN1; he&PNP;  
is&VBZ; amazed&AJ0-VVN; and&CJC;  
delighted&AJ0-VVN; to&TO0;  
be&VBI; offered&VVN; a&AT0; Huge&AJ0;  
Reward&NN1; for&PRP; what&DTQ;  
seems&VVZ; to&TO0; be&VBI;  
a&AT0;simple&AJ0; task&NN1;.&PUN;
```

# Corpora+POS

- most corpora nowadays are tagged at least with ‘part of speech’ information
- this can then be used in queries asked of the corpus
- POS-tagging for English is quite reliable



# Morphological analysis:

e.g. "These were only some simplest sample sentences."

These	these+Det+Pl
These	these+Pron+NomObl+3P+Pl
were	be+Verb+PastTense+Pl
only	only+Adj
only	only+Adv
only	only+Conj+Sub
some	some+Pron+NomObl+3P+Pl
some	some+Det+Sp
simplest	simple+Adj+Sup
sample	sample+Noun+Sg
sample	sample+Verb+Pres+Non3sg
sentences	sentence+Noun+Pl
sentences	sentence+Verb+Pres+3sg
.	.+?

Results from the  
Xerox  
morphological  
analyser and  
tagger

... a typical HLT result



# Step 2

- Corpora: from raw text to marked-up text
- Text encoding in general
- Corpora: from marked-up text to structured data



# Text Encoding Initiative (1995)

- a large effort by the Association for Computing and the Humanities, the Association for Literary and Linguistic Computing, and others.
- published guidelines for encoding electronic forms of documents for exchange and research
- based on SGML (an existing standard)
- attempts to make the structural details of text clear for archival of editions, contrasting editions, etc.



# Text Encoding Initiative: example

“Have you, miss? Well, for sure!”

A short time after she pursued, “I seed you go out with the master, but I didn't know you were gone to church to be wed”; and she basted away. John, when I turned to him, was grinning from ear to ear.



“Have you, miss? Well, for sure!”

A short time after she pursued, “I seed you go out with the master, but I didn't know you were gone to church to be wed”; and she basted away. John, when I turned to him, was grinning from ear to ear.

Original

< p >

< q > Have you, miss? Well,  
for sure! </ q > < / p >

< p > A short time after she  
pursued,  
< q > I seed you go out with  
the master, but I didn't  
know you were gone to  
church to be wed </ q >;  
and she basted away. John,  
when I turned to him, was  
grinning from ear to ear.  
< / p >

XCES-conform markup

“Have you, miss? Well, for sure!”

A short time after she pursued, “I seed you go out with the master, but I didn't know you were gone to church to be wed”; and she basted away. John, when I turned to him, was grinning from ear to ear.

↑  
Original

< p >  
< q > Have you, miss? Well,  
for sure! </ q > < / p >  
< p > A short time after she  
pursued,  
< q > I seed you go out with  
the master, but I didn't  
know you were gone to  
church to be wed </ q >;  
and she basted away. John,  
when I turned to him, was  
grinning from ear to ear.  
< / p >

XCES-conform markup

“Have you, miss? Well, for sure!”

A short time after she pursued, “I seed you go out with the master, but I didn't know you were gone to church to be wed”; and she basted away. John, when I turned to him, was grinning from ear to ear.

Original

< p >  
< q > Have you, miss? Well,  
for sure! </ q > </ p >  
< p > A short time after she  
pursued,  
< q > I seed you go out with  
the master, but I didn't  
know you were gone to  
church to be wed </ q >;  
and she basted away. John,  
when I turned to him, was  
grinning from ear to ear.  
< / p >

XCES-conform markup

“Have you, miss? Well, for sure!”

A short time after she pursued, “I seed you go out with the master, but I didn't know you were gone to church to be wed”; and she basted away. John, when I turned to him, was grinning from ear to ear.

↑  
Original

<p>  
<q>Have you, miss? Well,  
for sure!</q></p>  
<p>A short time after she  
pursued,  
<q>I seed you go out with  
the master, but I didn't  
know you were gone to  
church to be wed</q>;  
and she basted away. John,  
when I turned to him, was  
grinning from ear to ear.  
</p>

XCES-conform markup

# TEI base tag sets

- sets of standardized tags for encoding:
    - prose
    - verse
    - drama
    - transcriptions of speech
    - print dictionaries
    - terminological databases
- ... and many  
more extensions  
and details...



# Simple ‘TEI-conform’ examples: prose

```
<body>
<p>I fully appreciate Gen. Pope's splendid achievements
with their invaluable results; but you must know that
Major Generalships in the Regular Army, are not as
plenty as blackberries.
</p>
</body>
```



# TEI-conform examples: verse

```
<lg n=I>
<l>I Sing the progresse of a
    deathlesse soule ,</l>
<l>Whom Fate, with God made,
    but doth not controule ,</l>
<l>Plac'd in most shapes; all times
    before the law</l>
<l>Yoak'd us, and when, and since,
    in this I sing.</l>
<l>And the great world to his aged evening;</l>
<l>From infant morne, through manly noone I draw.</l>
<l>What the gold Chaldee , of silver Persian saw,</l>
<l>Greeke brass, or Roman iron, is in this one;</l>
<l>A worke t'out weare Seths pillars, bricke and stone,</l>
<l>And (holy writs excepted) made to yeeld to none,</l>
</lg>
```



# TEI-conform examples: prose and edition-specific information

<p>I wrote to Moor House and to Cambridge immediately, to say what I had done: fully explaining also why I had thus acted. Diana and <pb ed=ED1 n='475'> Mary approved the step unreservedly. Diana announced that she would <pb ed=ED2 n='485'>just give me time to get over the honeymoon, and then she would come and see me.

*This markup records the differing pagination of two editions*



# Motivation for adoption of SGML

- a standard already agreed upon in the print industry for re-use of content
- formal specification allows **validation** of documents marked up as TEI-conformant documents
- aspects of an interpretation of a document are explicitly represented and so can be used for indexing and retrieval



SGML documents must have a

# Document Type Definition

```
<act><title>Act I</title>
  <scene><title>Scene I. Elsinore. A platform before the castle.
    </title>
    <stagedir>FRANCISCO at his post. Enter to him BERNARDO.
  </stagedir>
  <speech>
    <speaker>BERNARDO</speaker>
    <line>Who's there?<line>
  </speech>
```

DTD:

```
<!ELEMENT play (title, personae, scndesc, playsubt,
  prologue?, act+, epilogue?)>
<!ELEMENT act (title, subtitle*, prologue?, scene+, epilogue?)>
<!ELEMENT scene (title, subtitle*, (speech | stagedir | subhead)+)>
<!ELEMENT speech (speaker+, (line | stagedir | subhead)+)>
```



# Step 3

- Corpora: from raw text to marked-up text
- Text encoding in general
- **Corpora:**  
**from marked-up text to structured data**



# Going beyond POS-tagging

- the more linguistic information that a corpus provides, the greater its utility:
  - searching for particular grammatical configurations is possible
  - using the information for training parsers is possible
  - evaluating linguistic accounts by larger-scale comparison of predicted and observed is encouraged



# The Penn Treebank (1994)

- 1 million words of newspaper text
- syntactically annotated

(TOP (S (NP-SBJ my best friend)  
      ( VP gave  
          ( NP me)  
          ( NP chocolate)  
          ( NP-TMP yesterday))  
      .))



# The Prague Dependency Treebank (1997)

- full morphological tagging
- syntactic analysis using dependency syntax  
(Panovová, Bémová)
- ‘tectogrammatical’ level (“linguistic meaning”: e.g., participant roles)
- initial goal: 200,000 sentences to be annotated



# The “International Corpus of English”

- Each ICE Corpus is divided into 2,000 word text samples representing various kinds of spoken and written English
- 500 texts: 200 written, 300 spoken
- the texts in ICE-GB were collected between 1990 and 1996
- A fully tagged and parsed corpus is only as useful as the tools that are provided to access it!

Greenbaum, Sidney (1988) ‘A Proposal for an International Corpus of English’,  
*World Englishes* 7: 315.

Nelson, Gerald (1996a) ‘The Design of the Corpus’, in S. Greenbaum (ed.),  
*Comparing English Worldwide: The International Corpus of English*, Oxford:  
Clarendon Press, 27-35

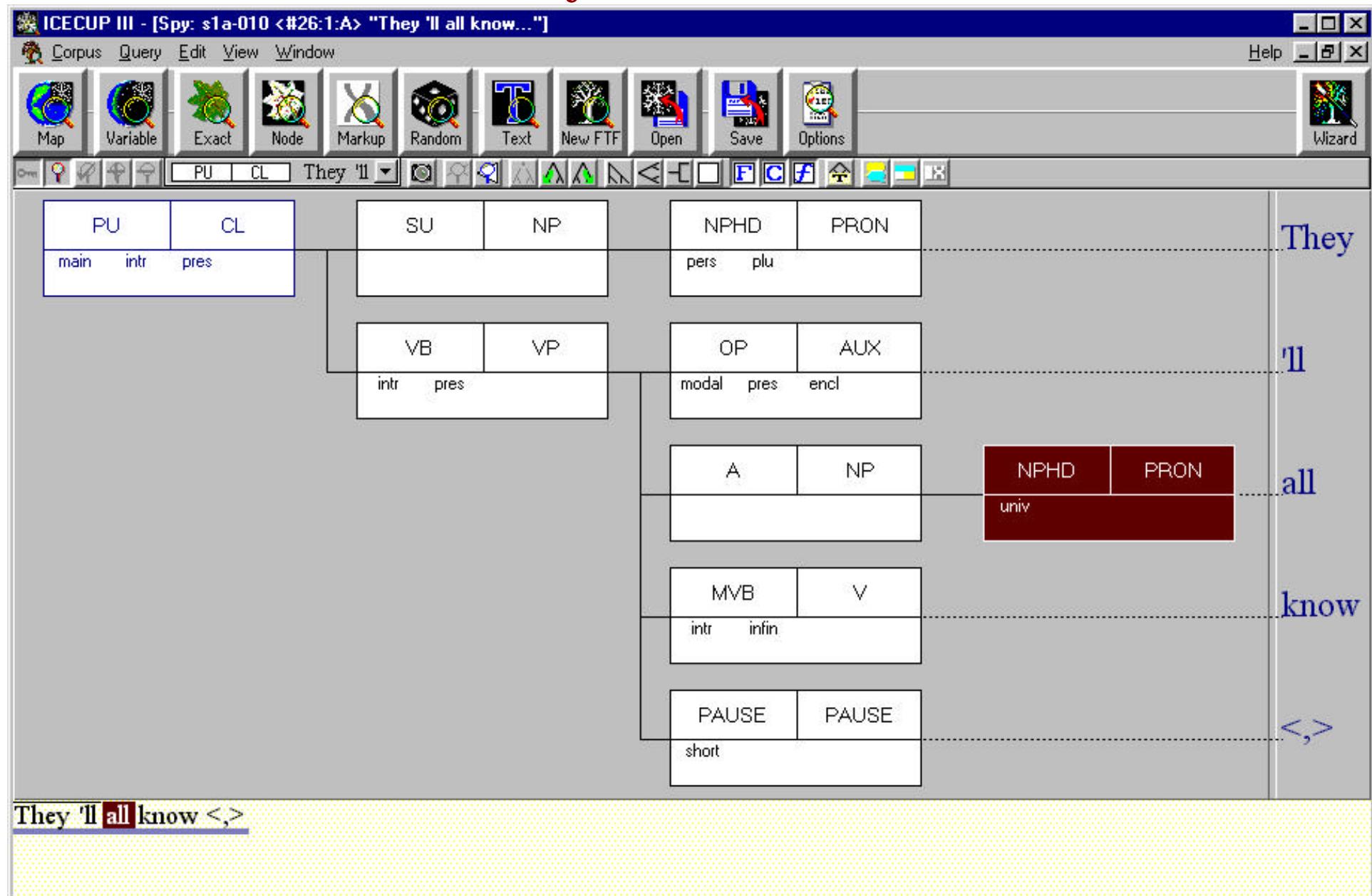
Nelson, Gerald (1996b) ‘Markup Systems’, in S. Greenbaum (ed.), (*op.cit.*), pp36-53



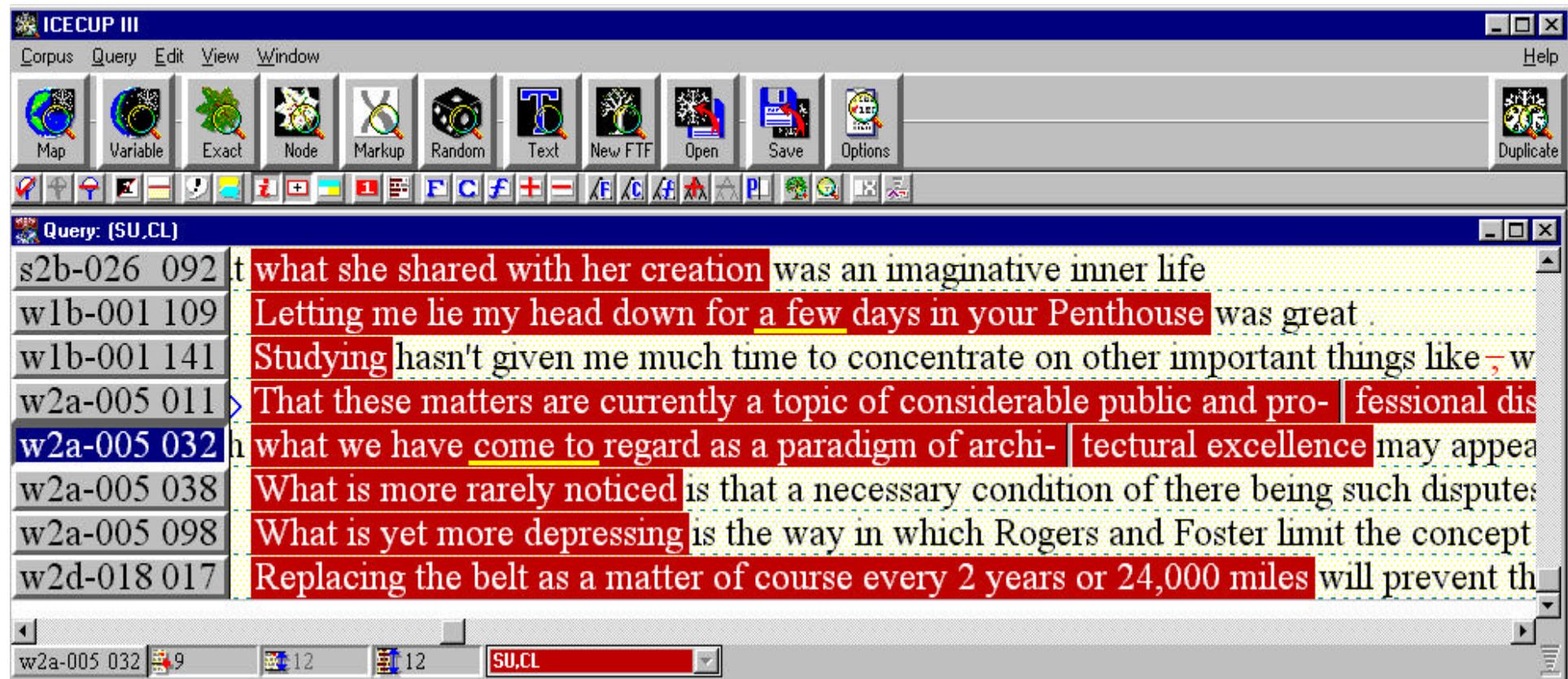
Query Result ‘all’ : International Corpus of English (concordanced)

		I mean I didn't know Islington at all until I moved there really
s1a-010 024		They'll all know <,,>
s1a-010 026		They're all young and very wet behind the ears <,,>
s1a-010 028		Have you got all the vowels <,,>
s1a-010 074		Oh all the same is it <,,>
s1a-010 084		First of all you don't score so much
s1a-010 161		all the same
s1a-010 176		kids she goes back to Felicity and all her achievements <,,>
s1a-010 211		... ↳ he immediately uh was uhm all over her or som ... ↳ I don't know
s1a-010 229		ow ten O levels and getting them all and then going on to A levels and doing th
s1a-010 239		Done all right <,,>
s1a-010 243		t I'm trying to say that you know all these things that Linda sets such great sto
s1a-010 247		these people who was really into all the trappings of power <,,>
s1a-010 253		now it'd be his private plane and all that nonsense
s1a-010 257		wers weren't the right colour and all that nonsense you know and sort of <,,> bi
s1a-010 258		I mean all that and he's ↳ he's popped his clogs at 1
s1a-010 262		

# Structural Analysis of selected Sentence



# Search Results for Tree Fragment: *Subject filled by Clause*



The screenshot shows the ICECUP III software interface. The window title is "ICECUP III". The menu bar includes Corpus, Query, Edit, View, Window, and Help. The toolbar contains icons for Map, Variable, Exact, Node, Markup, Random, Text, New FTF, Open, Save, Options, and Duplicate. Below the toolbar is a row of smaller icons. The main window displays a list of search results in a table format. The columns are labeled with file names and line numbers. The text in the results is highlighted in red, indicating matches for the search query. The results show various clauses where the subject is filled by a clause.

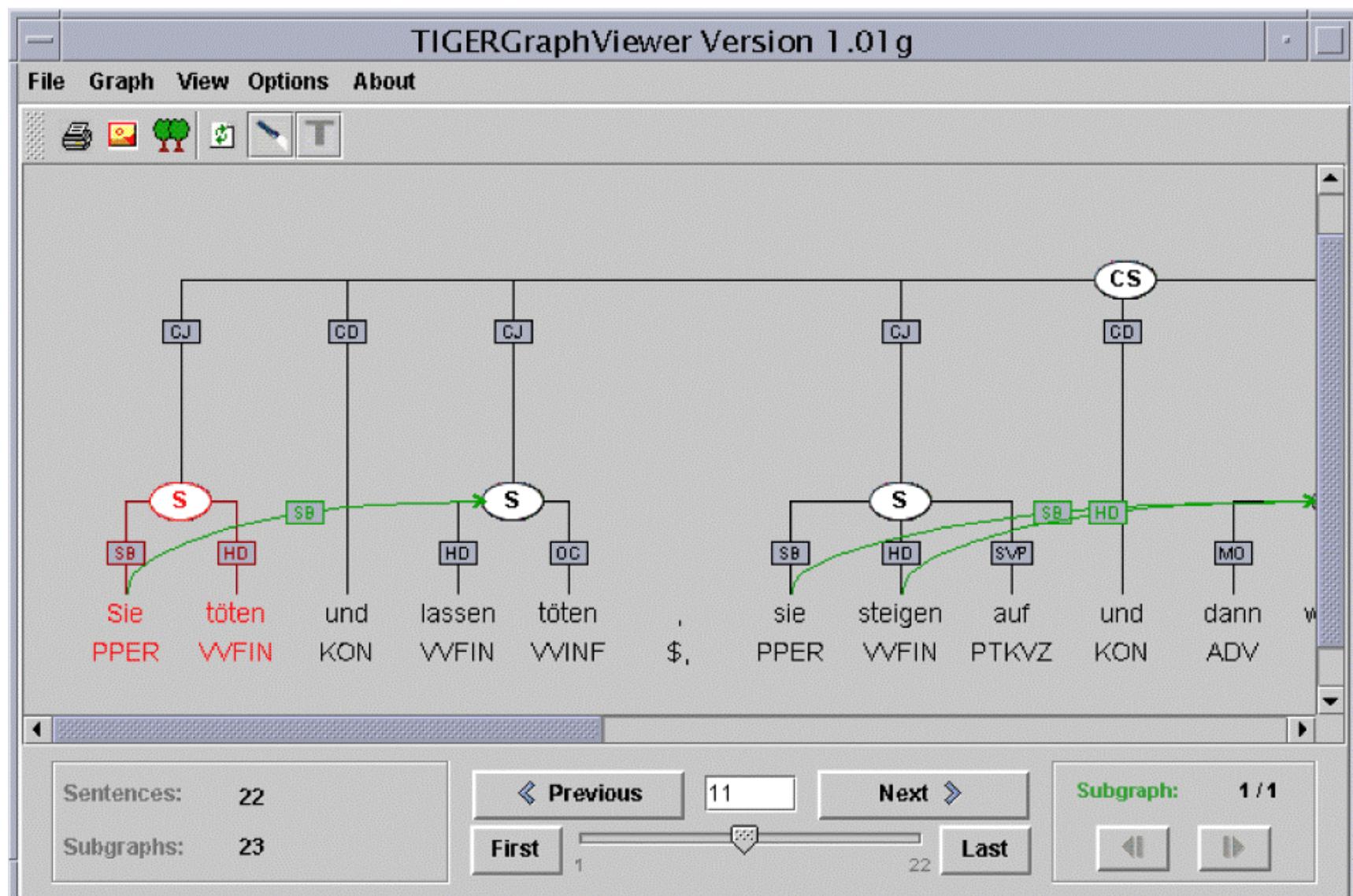
		Query: (SU,CL)
s2b-026	092	t what she shared with her creation was an imaginative inner life
w1b-001	109	Letting me lie my head down for a few days in your Penthouse was great .
w1b-001	141	Studying hasn't given me much time to concentrate on other important things like , w
w2a-005	011	> That these matters are currently a topic of considerable public and pro- fessional dis
w2a-005	032	what we have come to regard as a paradigm of archi- tectural excellence may appear
w2a-005	038	What is more rarely noticed is that a necessary condition of there being such disputes
w2a-005	098	What is yet more depressing is the way in which Rogers and Foster limit the concept
w2d-018	017	Replacing the belt as a matter of course every 2 years or 24,000 miles will prevent th



# The TIGER treebank (2003)

- German
- 35,000 newspaper sentences
- target 80,000 sentences by end of project
- automatic parsing of corpus (broad coverage LFG)
- conversion of parse results into more ‘neutral’ form





# Other treebanks / parsed corpora

- Susanne corpus (Sampson, 1995)
- Lancaster parsed corpus (Leech, 1992)
- Under development:
  - French
  - Spanish
  - Italian
  - Bulgarian
  - Russian



# ...three steps.

- Corpora: from raw text to marked-up text
- Text encoding in general
- Corpora: from marked-up text to structured data



# Several problems and a solution

- how to overcome the one corpus - one tool syndrome?
- how to move between different representations of similar information?
- how to increase the complexity and breadth of linguistic annotation?



# Solution: XML and related technology

- XML - the extensible markup language - replaces SGML as the markup language of choice
- strongly supported by software developers (W3C, Java, ...)
- advanced tools becoming freely available:  
including **multilayer annotation**
- already the representation language of choice for corpora



# Current Corpus Annotation Standards

Text Encoding Initiative  
(TEI)

Corpus Encoding Standard

Generalized Markup  
and Tools  
(XML)

XCES



# Problem of intersecting hierarchies

- XML allows only ‘balanced’ bracketting
- Brackets may not cross each other
- But: many kinds of information cannot be combined into single hierarchies...
- This is a very common problem: also well-known to us in linguistics



# Basic ‘stand-off’ annotation

## XML ‘base’ document

```
<w id="u-01">Have</w>
<w id="u-02">you</w>
<punc type="comma" id="u-03">,</punc>
<w id="u-04">miss</w>
<punc type="question" id="u-05">?</punc>
<w id="u-06">Well</w>
<punc type="comma" id="u-07">,</punc>
<w id="u-08">for</w>
<w id="u-09">sure</w>
<punc type="exclamation"
      id="u-10">!</punc>
```

## XML document for page breaks

```
<page id="page-01" from="..." to="u-07"/>
<page id="page-02" from="u-08" to="..."/>
```

## XML document for sentences

```
<s id="s-01" from="u-01" to="u-05"/>
<s id="s-02" from="u-06" to="u-10"/>
```



# Technology developments

- XML is set to replace HTML as the basic language of the World-Wide Web
- XML extensions provide increasing functionality
  - translations between XML schemes
  - transparent interfaces between XML and DB
  - flexible rendering: graphical, typesetting, ...
- the technology is already moving inside Web-browsers...



**Name des Bereiches**

---

Fachbereich 00  
**Name des Fachbereiches**

---

# Implications



Universität Bremen

Modes:

Segment

Edit

Segment

Sentences

Paragraphs

Segment At

Other:

Color Code

Search:

Find

The story was about a boy who had a bucket a net and a dog and the dog took the bucket and the boy took the net and they walked over to go to the pond to catch a frog but when they went they looked all over almost all day but they couldn't find the frog.

So they went and looked over in a tree. They didn't find anything but then they looked in a pond and they saw a frog on a lily pad and then they noticed that the frog wasn't looking so they jumped at him but when they jumped they accidentally tripped over a log and fell into the water and then they went over and then the boy was looking straight at the frog and the frog looked at him and the frog was getting angry for being chased so they tried to chase him okay so they chased him around and he hopped onto a log and the told the dog to go to one end and the boy went to the other so that the frog couldn't get away and the dog started running towards the frog.

The frog turned away and the boy took his net and then he caught the dog by accident so he sat there looking at the frog and shaking his fist and the frog was lying on a rock and then he shook his fist again and said goodbye and then he ran home and he had a bath and the frog didn't know where the boy went and the frog said to himself "hey I don't have any friends. Maybe I should go with the boy so he goes with the boy and then the frog follows the tracks so he can get to the boy's house then he gets there and he finds himself into a big house and the frog follows the footprints and he goes under the door and then he looks at the dog and the boy and they're having a bath so the frog hopped in and they play together and then the frog hopped on the dog's head. ||

Mick O'Donnell's systemic coder  
(wagsoft.com)

Zoom %

100

Layout Version

Normal Graph

Actions

ReDraw

Print Scheme

Find Feature

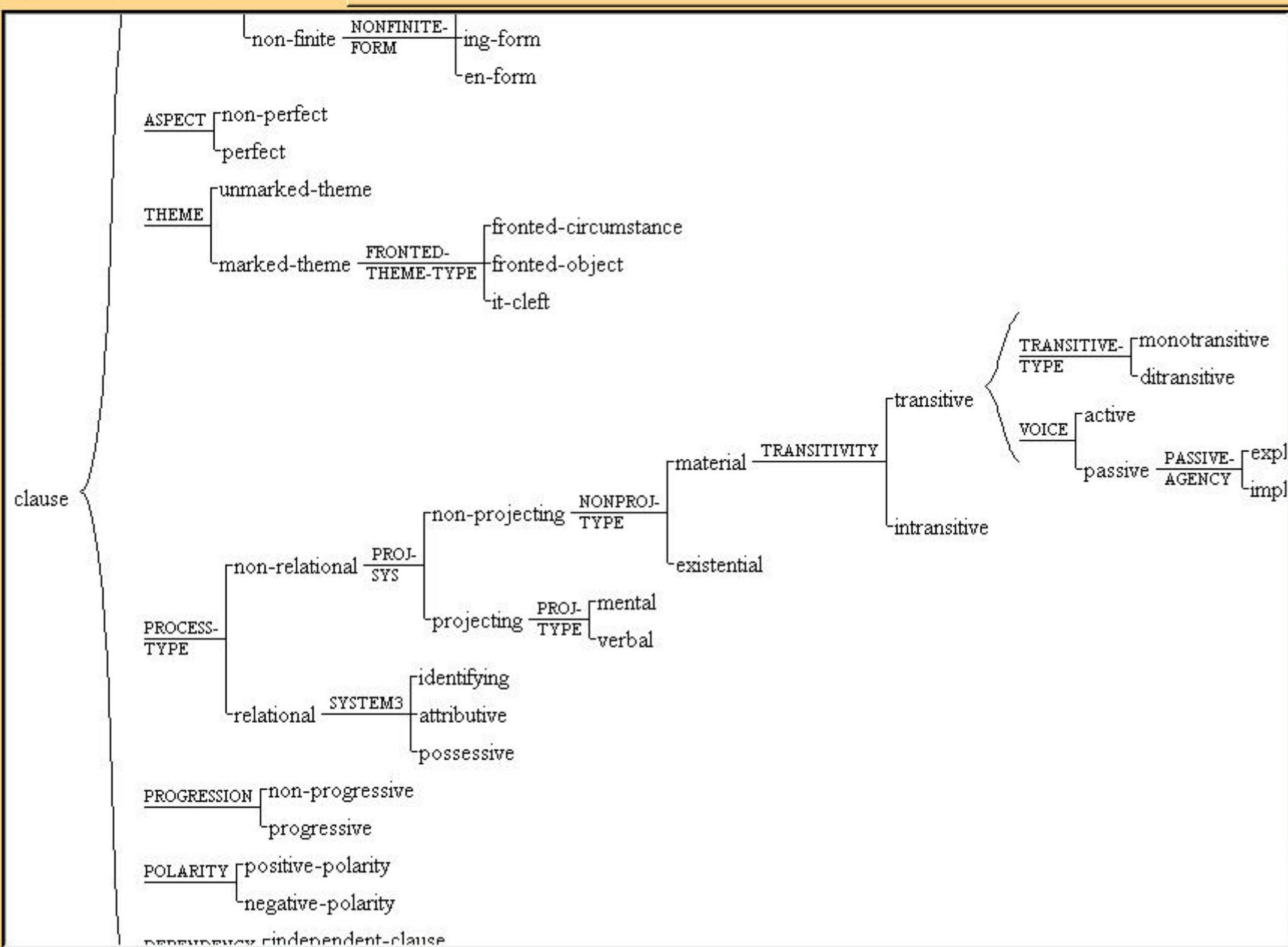
Find System

StartFeat

clause

Depth

all



Codings Scheme Options Help

Text Scheme Coding Review Statistics Codings File: D:/Systemic-tools/MickCoder463/Coder463/Texts/330aFrog-Text.cd3  
Scheme File: D:/Systemic-tools/MickCoder463/Coder463/Schemes/grammar.scheme

Define and modify the Coding Scheme

Item 1 of 2

Text: The story was about a boy who had a bucket a net and a dog

Goto

Previous

Next

Next Incomplete

First

Actions

Ignore

Text:

Comment:

Select

Choose from the following, then hit 'Select'.

Selected Features clause

POLARITY

positive-polarity  
 negative-polarity

DEPENDENCY

independent-clause  
 dependent-clause

SECTION

introduction  
 contents  
 abstract

FINITENESS

finite  
 non-finite

ASPECT

non-perfect  
 perfect

THEME

unmarked-theme  
 marked-theme

PROCESS-TYPE

non-relational  
 relational

PROGRESSION

non-progressive  
 progressive

# Interoperability problems

- coding scheme may only be changed with the tool itself
- use of coding schemes from elsewhere then not possible
- use of coding scheme elsewhere not possible



# Kay O'Halloran: Systemics

(University of Singapore)

- Very nice tool
- Easy to prepare complex multilayered functional analyses
- Covers conjunctive relation, exchange structure, and cohesion analyses
- But: what do you do with the results?
- Technology limited (printing methods)



76 Systemics 1.0

File Pages

Text Clause Interclausal Discourse Search Grammar About

Clause : 1

The antarctic explorer felt the wind and snow blowing against her face.

TH1	Topic Theme										
M1	Subj		Fin Temp								
	Mood		Mood	Residue							
T1	Sens		Proc Ment	Phen Range							
			Perc								
E1	Agent										
LEX1											
MET1											

Tex1 Topic unmarked  
---

Int1 Mood declarative full

Lex1

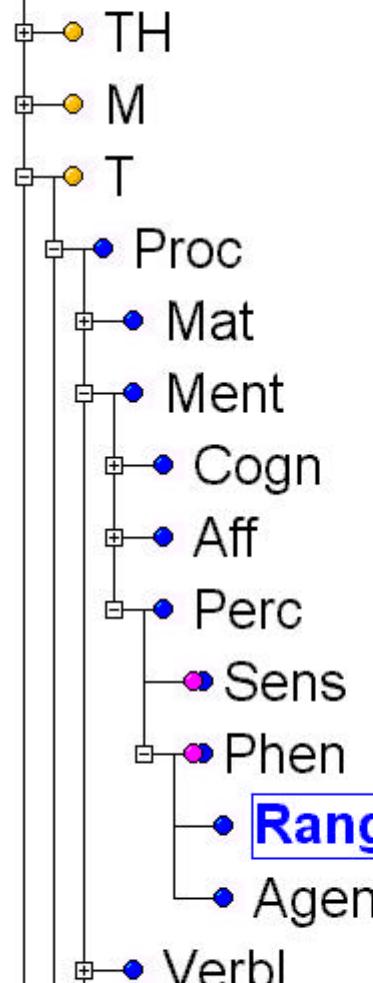
Exp1 VOICE

Preparing clause ... Done.

## System tree:

File Edit View

## Clause (CI)



## Glossary

Background colour for label

(CI/T/Proc/Ment/Perc) allows using (CI/T/Phen/Range)

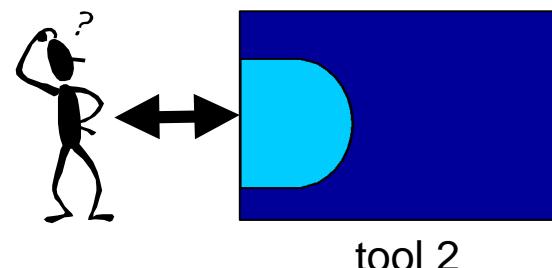
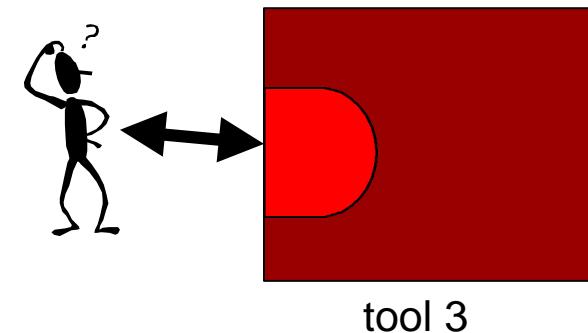
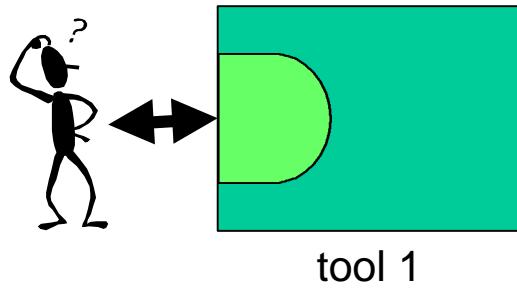
Phen/Range = Phenomenon/Range is that which is being sensed by the Senser. This may be realised as a participant in the clause, a rankshifted clause (if the event is a single complex phenomenon) or a projected clause (if the event is a complex phenomenon of a different order of reality).

# Interoperability problems

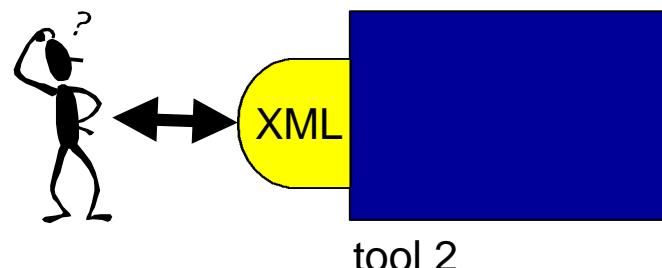
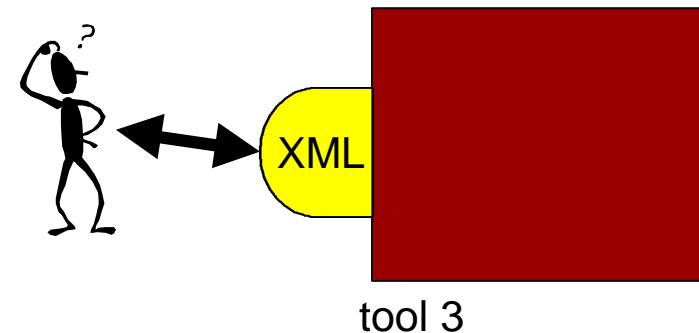
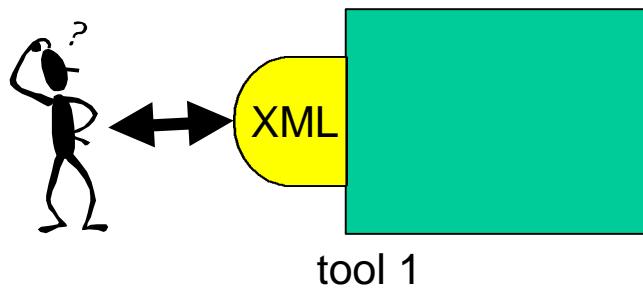
- no standardized output forms for analyses
- mixture of display and content
- no standardized output forms for grammar
- mixture of display and content
- coding scheme may only be changed with the tool itself



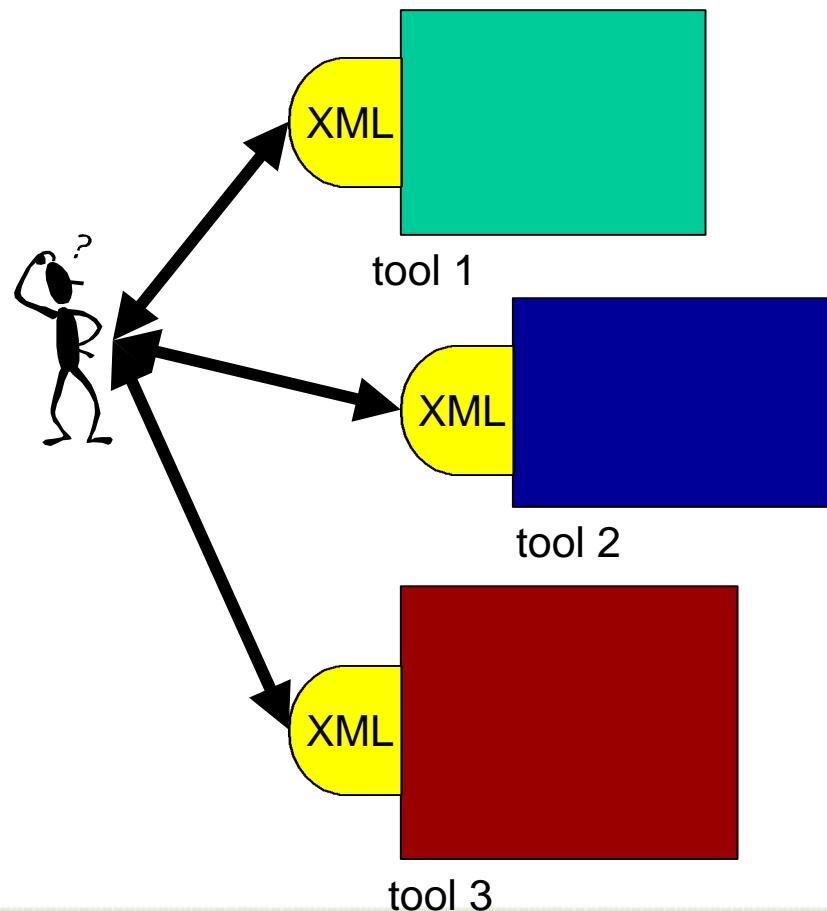
# 1990s view of linguistic tools



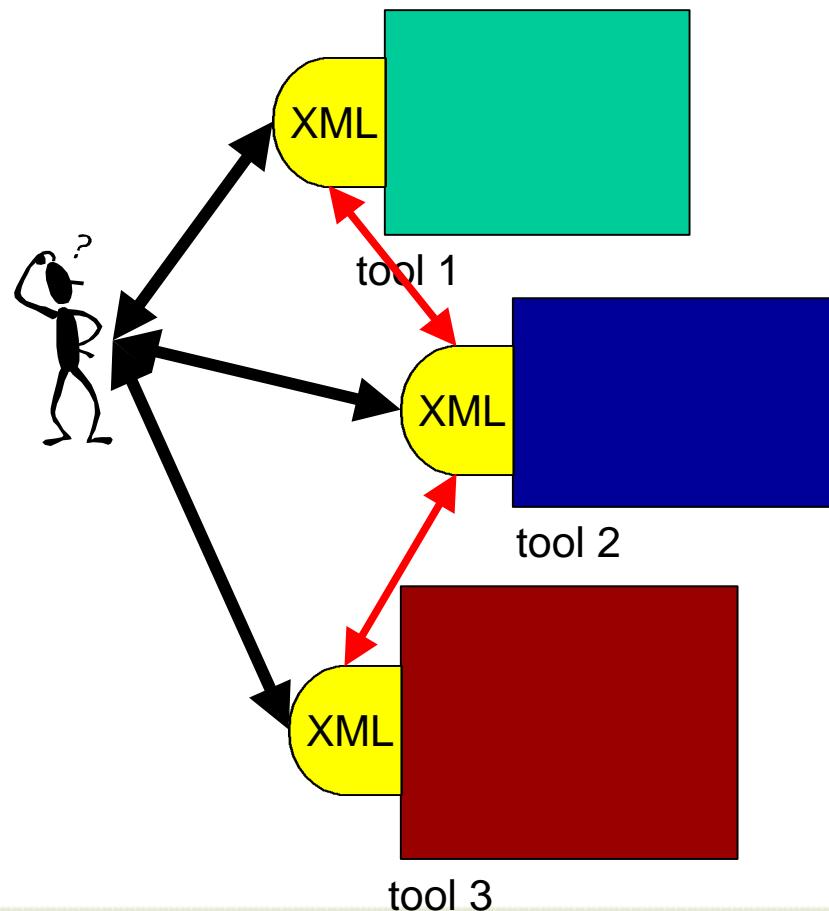
# 2000s view of linguistic tools



# 2000s view of linguistic tools



# 2000s view of linguistic tools



# Current project

- to provide XML schema definitions for the main theoretical constructs in SFL
  - system networks
  - realization statements
  - instantiated syntagmatic structures
- to provide XML wrappers around existing tools to improve their interoperability



Name des Bereiches

---

Fachbereich 00  
Name des Fachbereiches

---

# Example 1

Grammar debugging with KPML  
and coding with Mick's coder

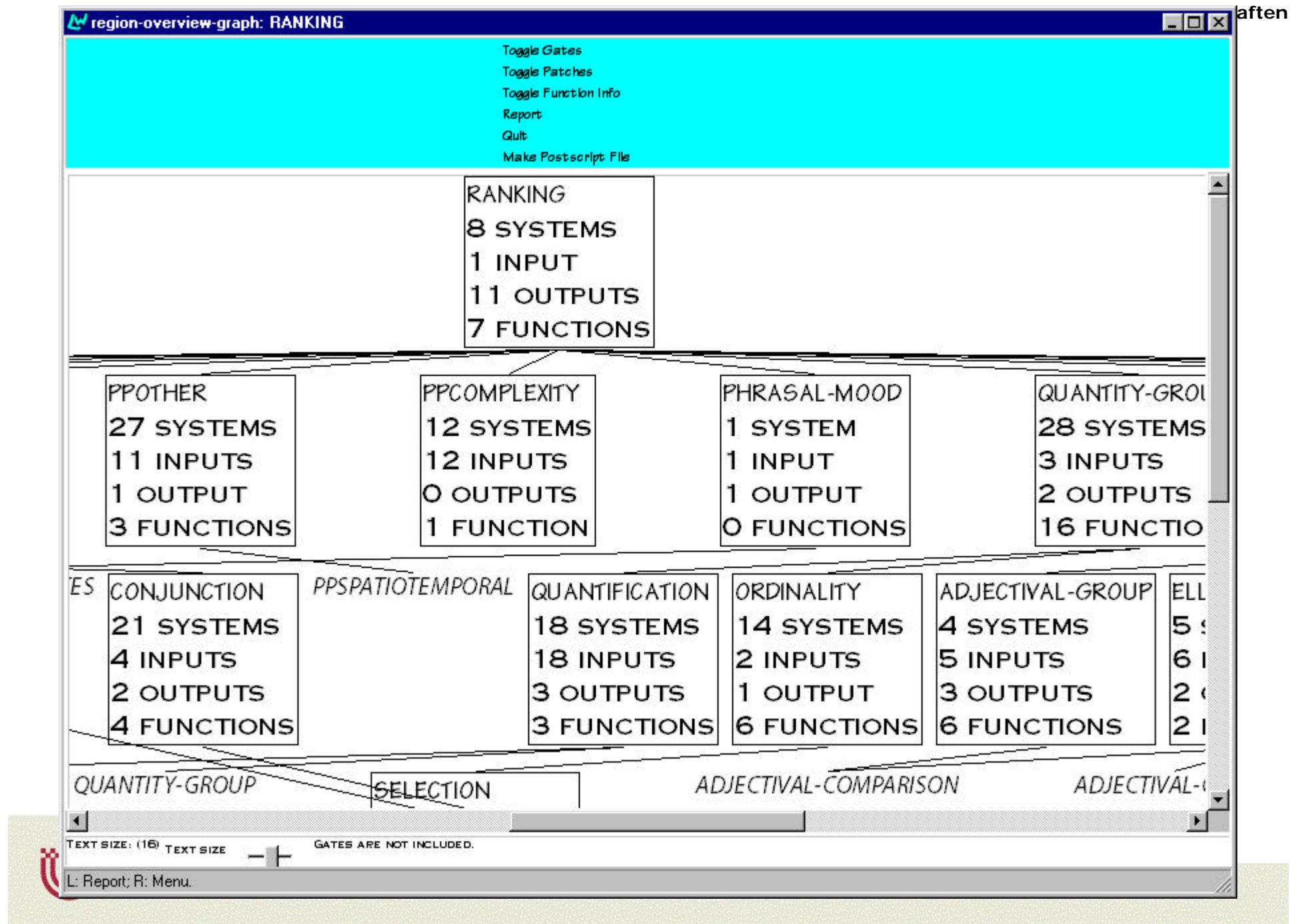


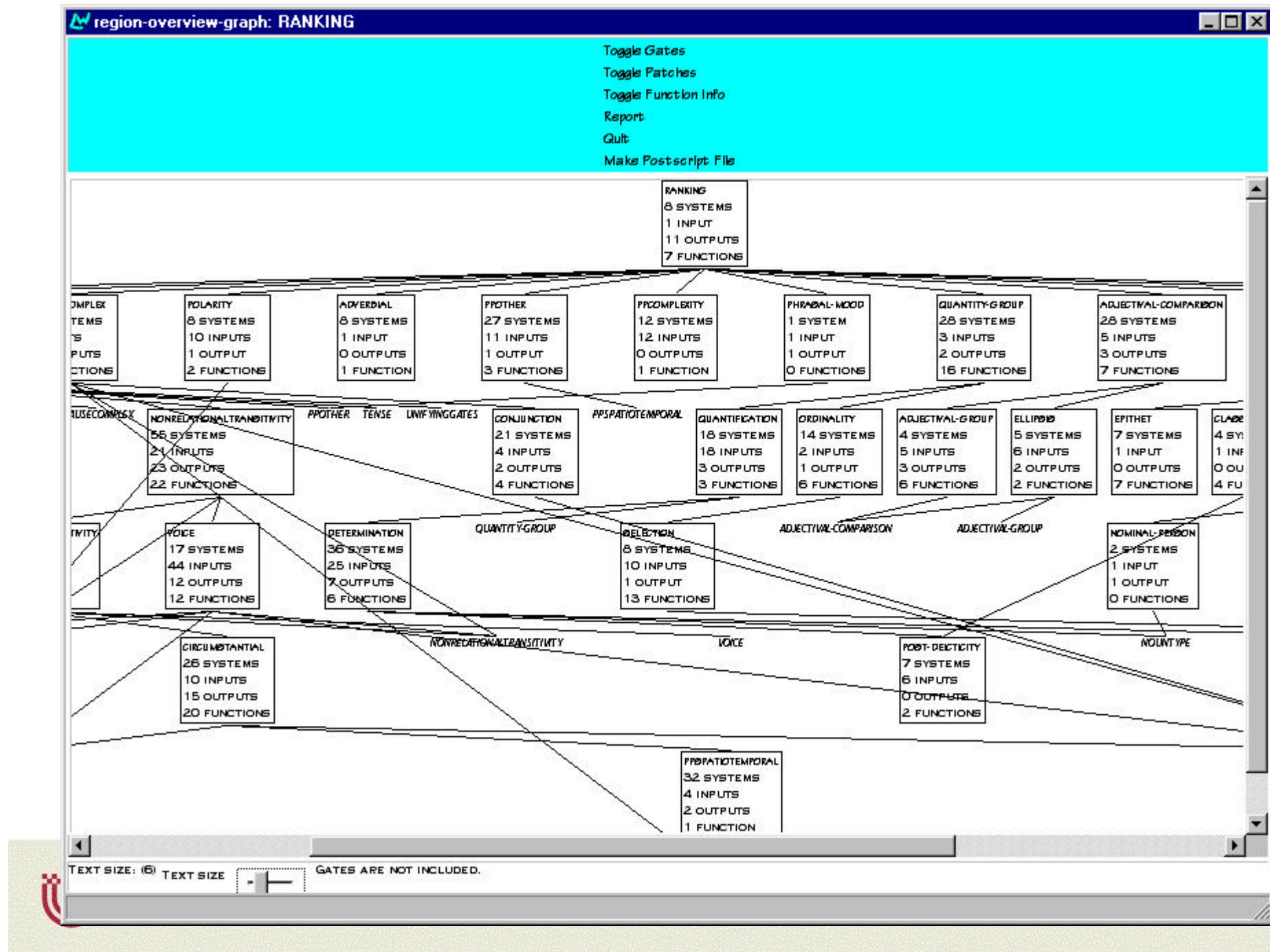
Universität Bremen

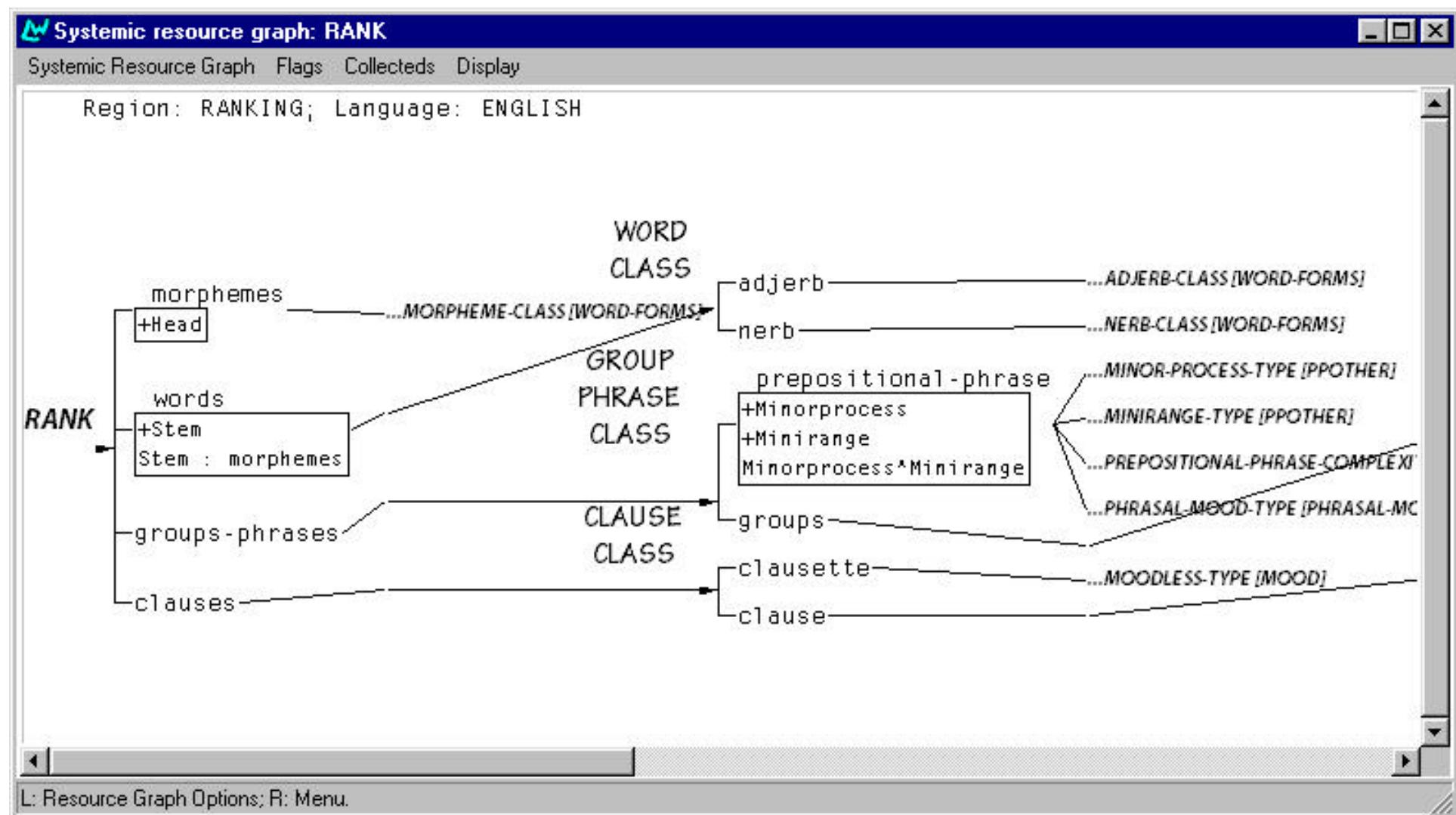
# KPML

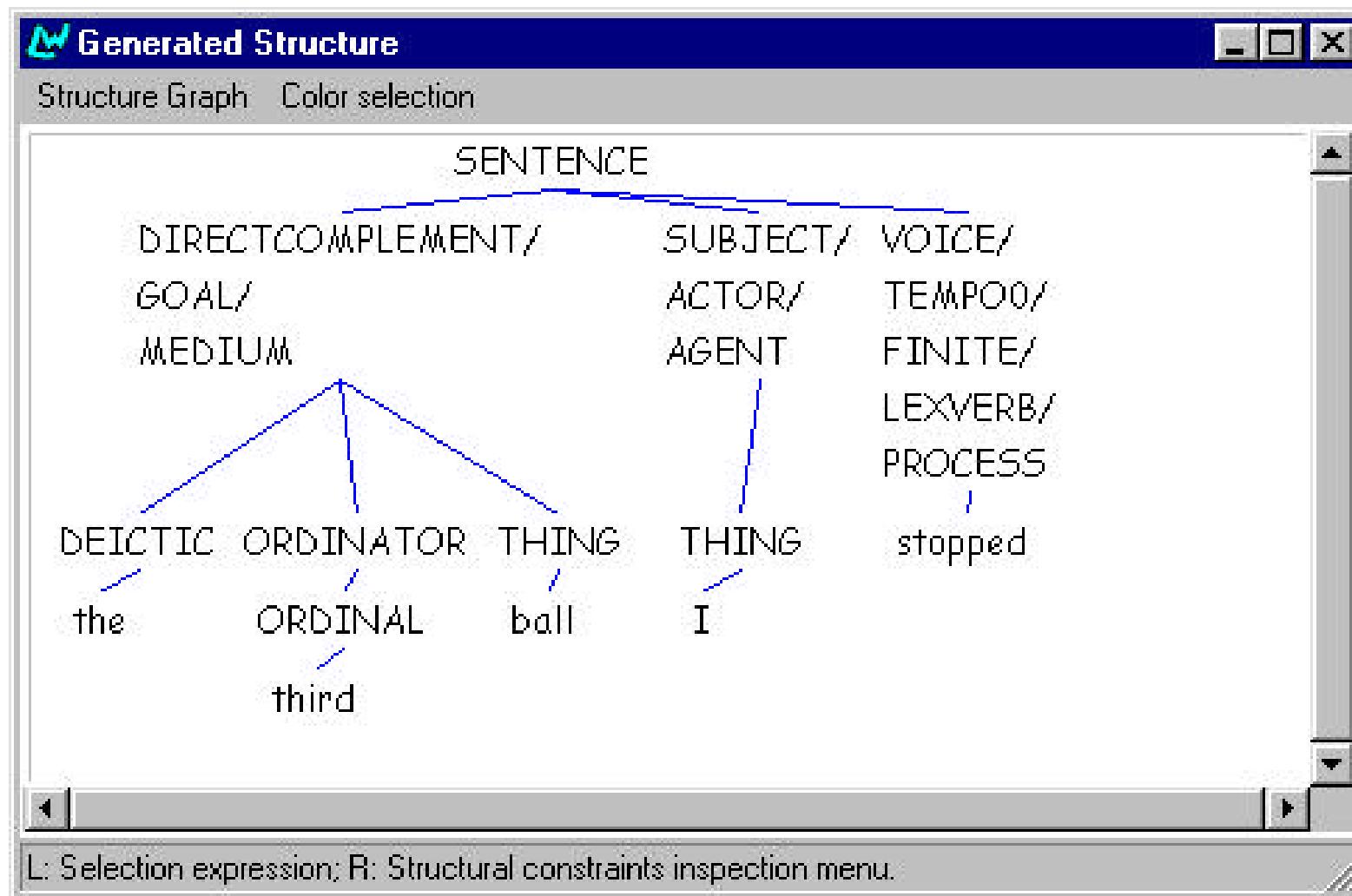
- a graphical development environment for large-scale systemic grammars built on top of Penman
- allows views of resources and their instantiations
- allows views according to axis, rank, metafunction, functional region and stratum
- strongly multilingual

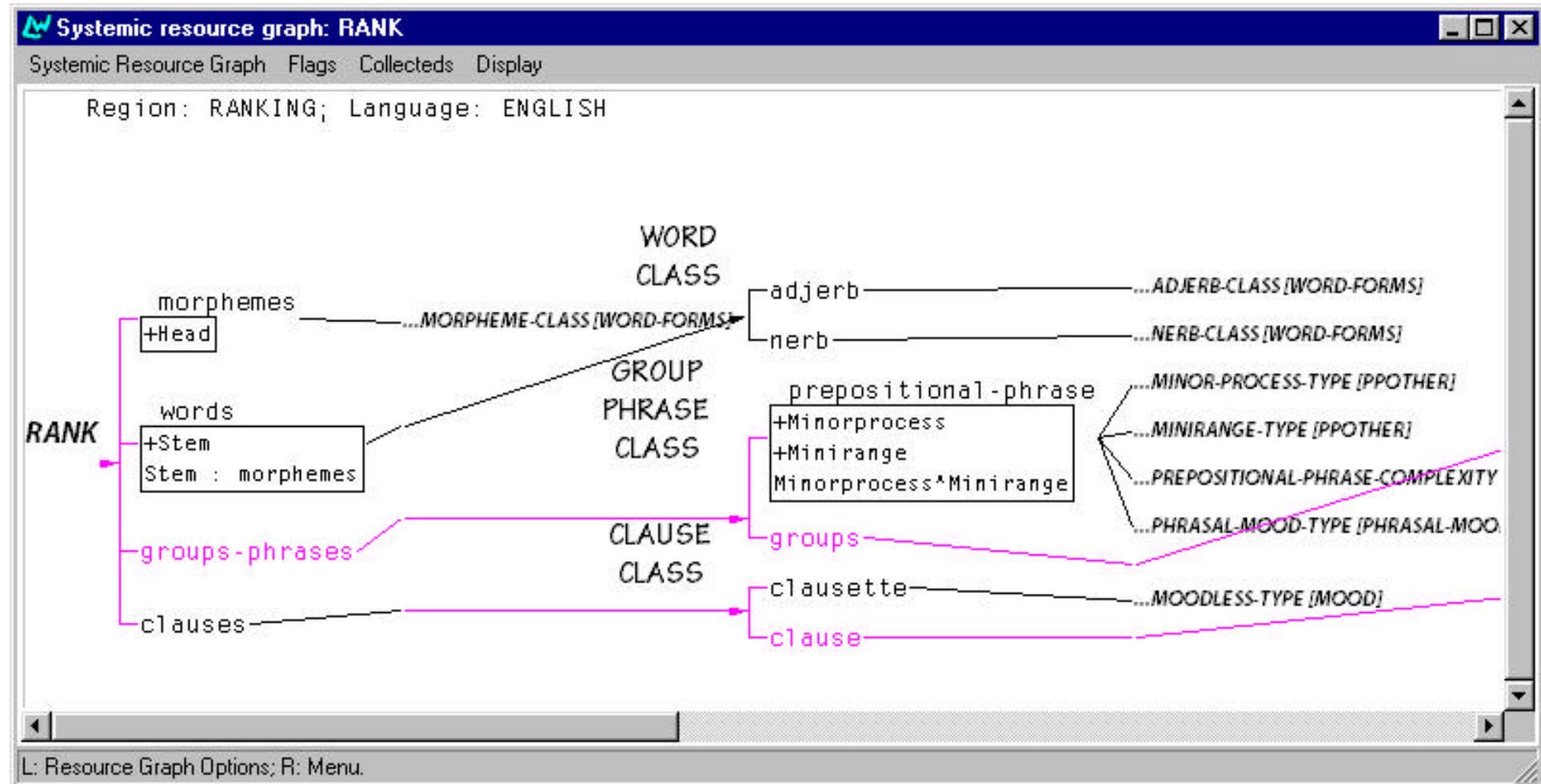








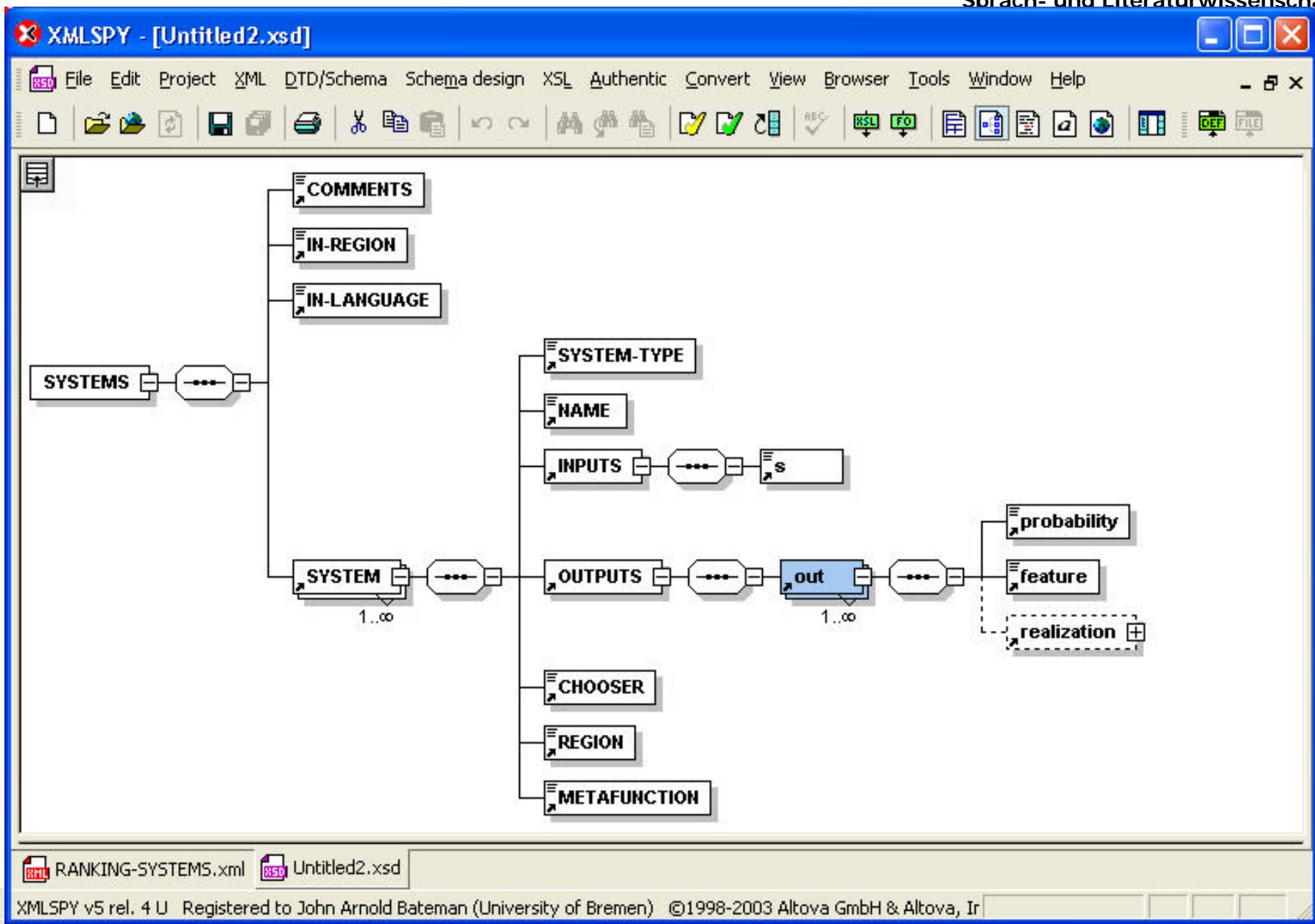




# Interoperability

- XML versions of systems and structures may be both accepted as input and produced as output





```
<SYSTEM>
  <SYSTEM-TYPE>SYSTEM</SYSTEM-TYPE>
  <NAME>CLAUSE-CLASS</NAME>
  <INPUTS><s>CLAUSES</s></INPUTS>
  <OUTPUTS>
    <out>
      <probability>0.5</probability>
      <feature>CLAUSE</feature>
    </out>
    <out>
      <probability>0.5</probability>
      <feature>CLAUSETTE</feature>
    </out>
  </OUTPUTS>
  <CHOOSER>CLAUSE-CLASS-CHOOSER</CHOOSER>
  <REGION>RANKING</REGION>
  <METAFUNCTION>LOGICAL</METAFUNCTION>
</SYSTEM>
```



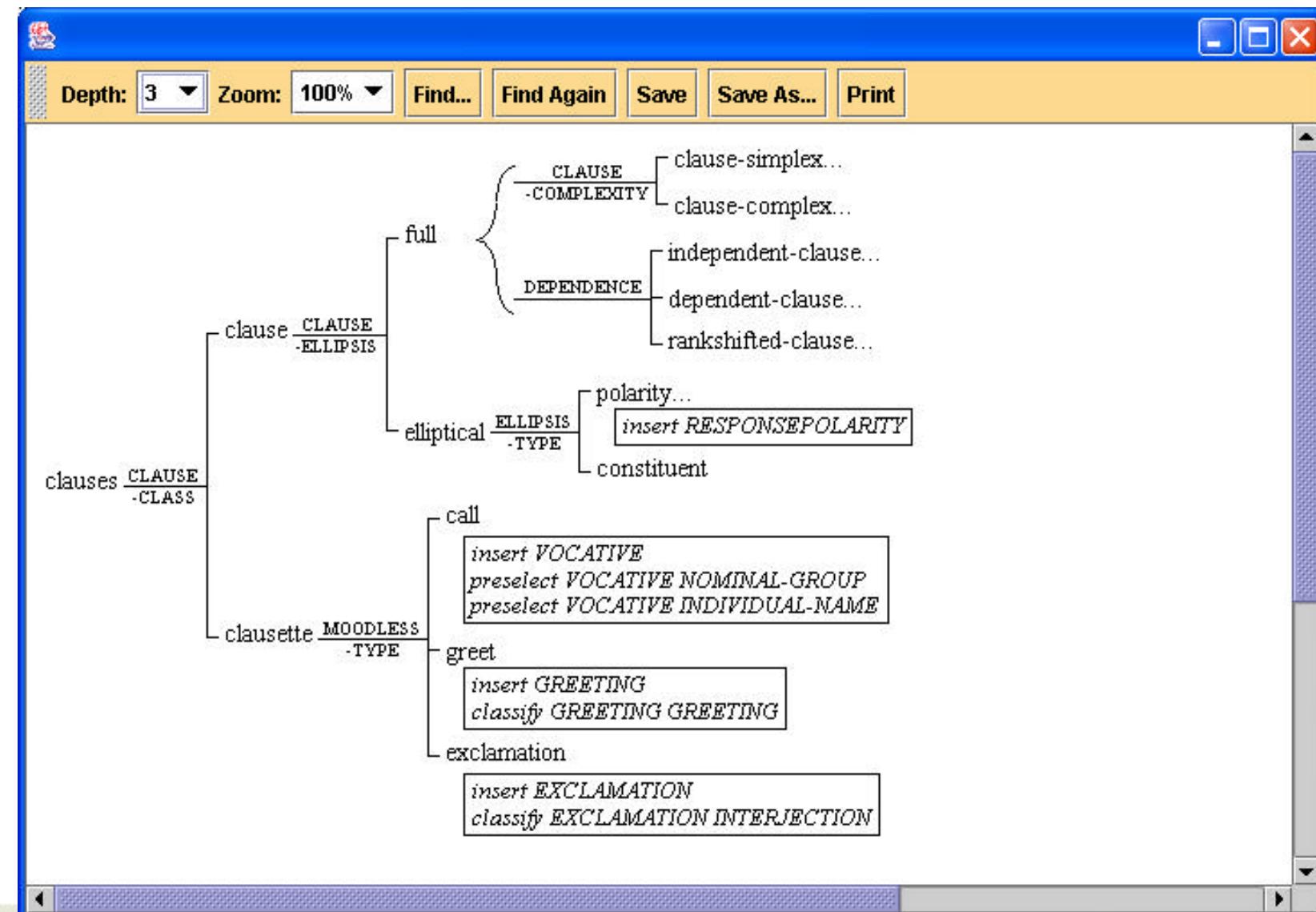
The screenshot shows a Windows application window titled "E:\systems\kpml\mick-grapher-trials\KPMILDemo\Grammar\RANKING-SYSTEMS.xml". The window has a standard Windows title bar with icons for minimize, maximize, and close. Below the title bar is a menu bar with "Datei", "Bearbeiten", "Ansicht", "Favoriten", "Extras", and "?". The main area contains a toolbar with icons for back, forward, search, and other file operations. The address bar shows the file path "E:\systems\kpml\mick-grapher-trials\KPMILDemo\Grammar\RANKING-SYSTEMS.xml". To the right of the address bar is a "Wechsel zu" button. The content area displays the following XML code:

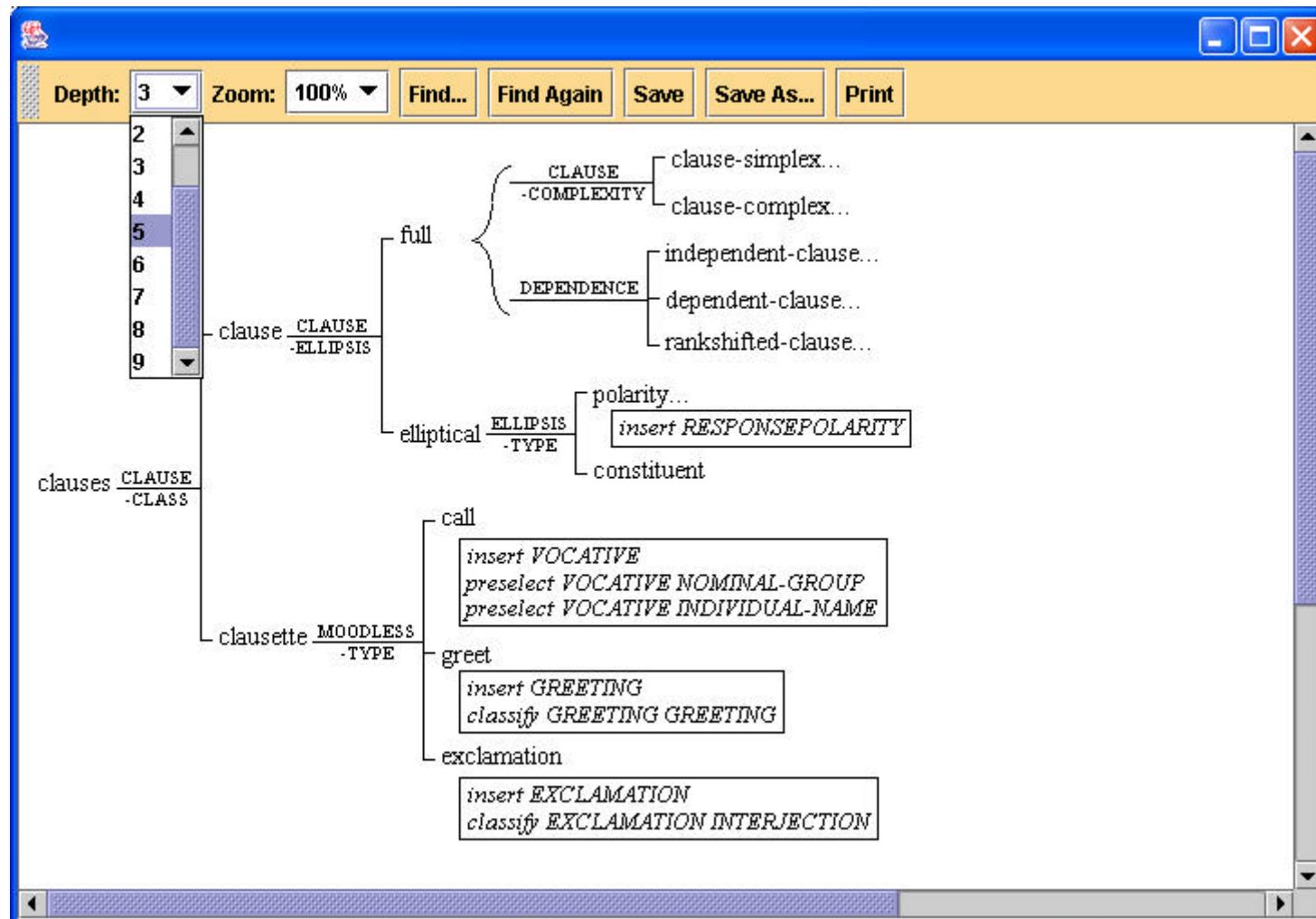
```
<IN-REGION>RANKING</IN-REGION>
<IN-LANGUAGE>ENGLISH</IN-LANGUAGE>
- <SYSTEM>
  <SYSTEM-TYPE>SYSTEM</SYSTEM-TYPE>
  <NAME>CLAUSE-CLASS</NAME>
  - <INPUTS>
    <s>CLUSES</s>
  </INPUTS>
  - <OUTPUTS>
    - <out>
      <probability>0.5</probability>
      <feature>CLAUSE</feature>
    </out>
    - <out>
      <probability>0.5</probability>
      <feature>CLAUSSETTE</feature>
    </out>
  </OUTPUTS>
  <CHOOSER>CLAUSE-CLASS-CHOOSER</CHOOSER>
  <REGION>RANKING</REGION>
  <METAFUNCTION>LOGICAL</METAFUNCTION>
</SYSTEM>
+ <SYSTEM>
+ <SYSTEM>
+ <SYSTEM>
```

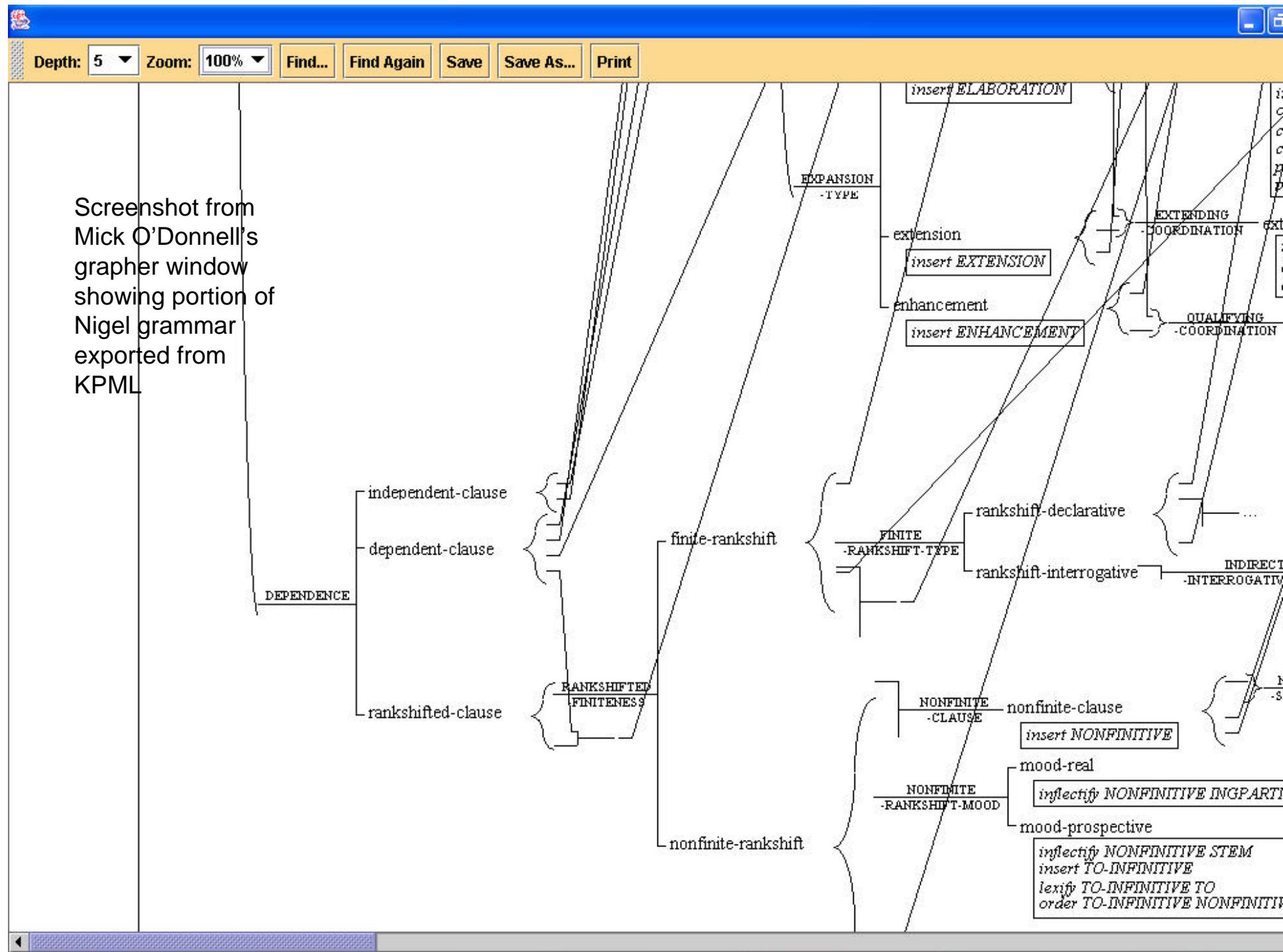
# Mick's coder

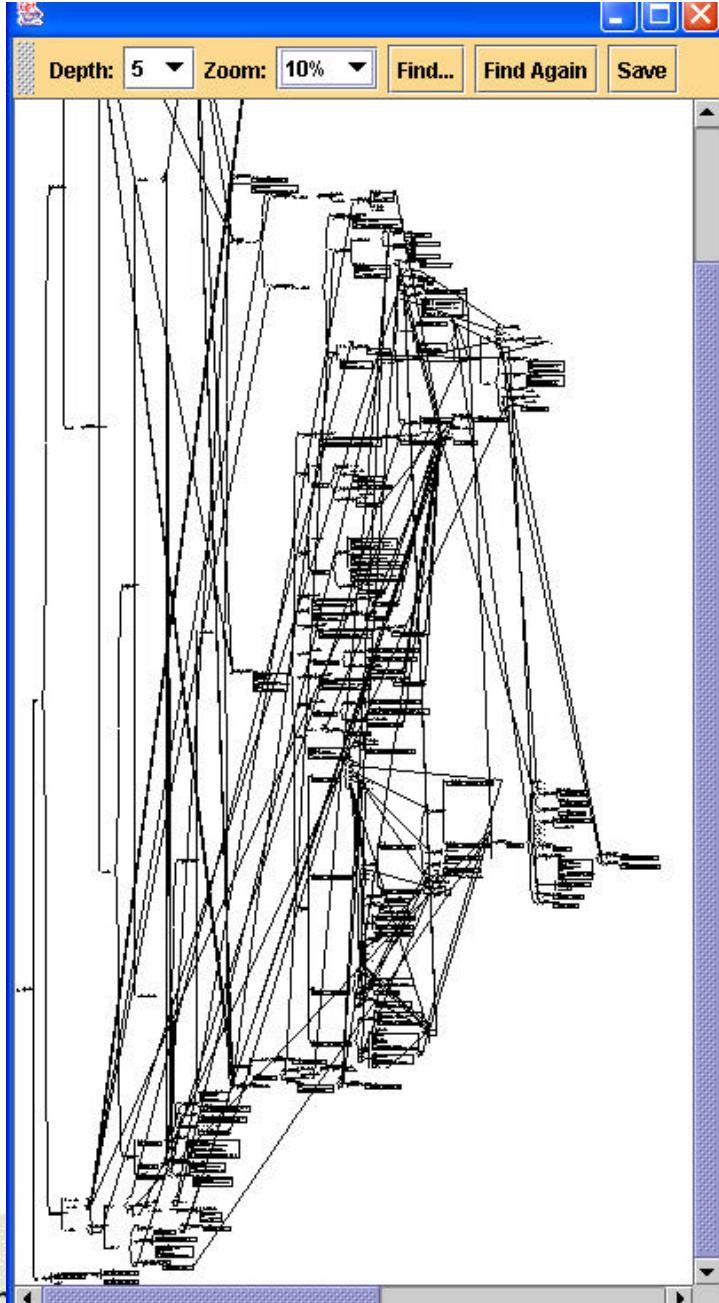
- extended to accept and produce the XML-definition of system networks











Interoperability is thus achieved between the two tools

## Example 2

Storing and viewing results of  
systemic analyses in various  
forms



# PoW Treebank

- Polytechnic of Wales parsed corpus (Fawcett, 1989)
- Child language corpus
- 65,000 words
- 11,000 trees
- Available through the International Computer Archive of Modern English (ICAME, Bergen)



# Example entry on ICAME CD

Z 1 CL 2 S NGP HP I 2 M PLAY 2 C PGP 3 P WITH [FS:MY-  
CHIP] 3 CV NGP 4 DD MY [RP:MY] 4 MO QQGP AX BIG 4 H  
TIPPER-LORRY 1 CL 5 & AND 5 S NGP HP I [RP:I] 5 M  
CALL 5 C PGP 6 PM FOR 6 CV NGP HN DAVID

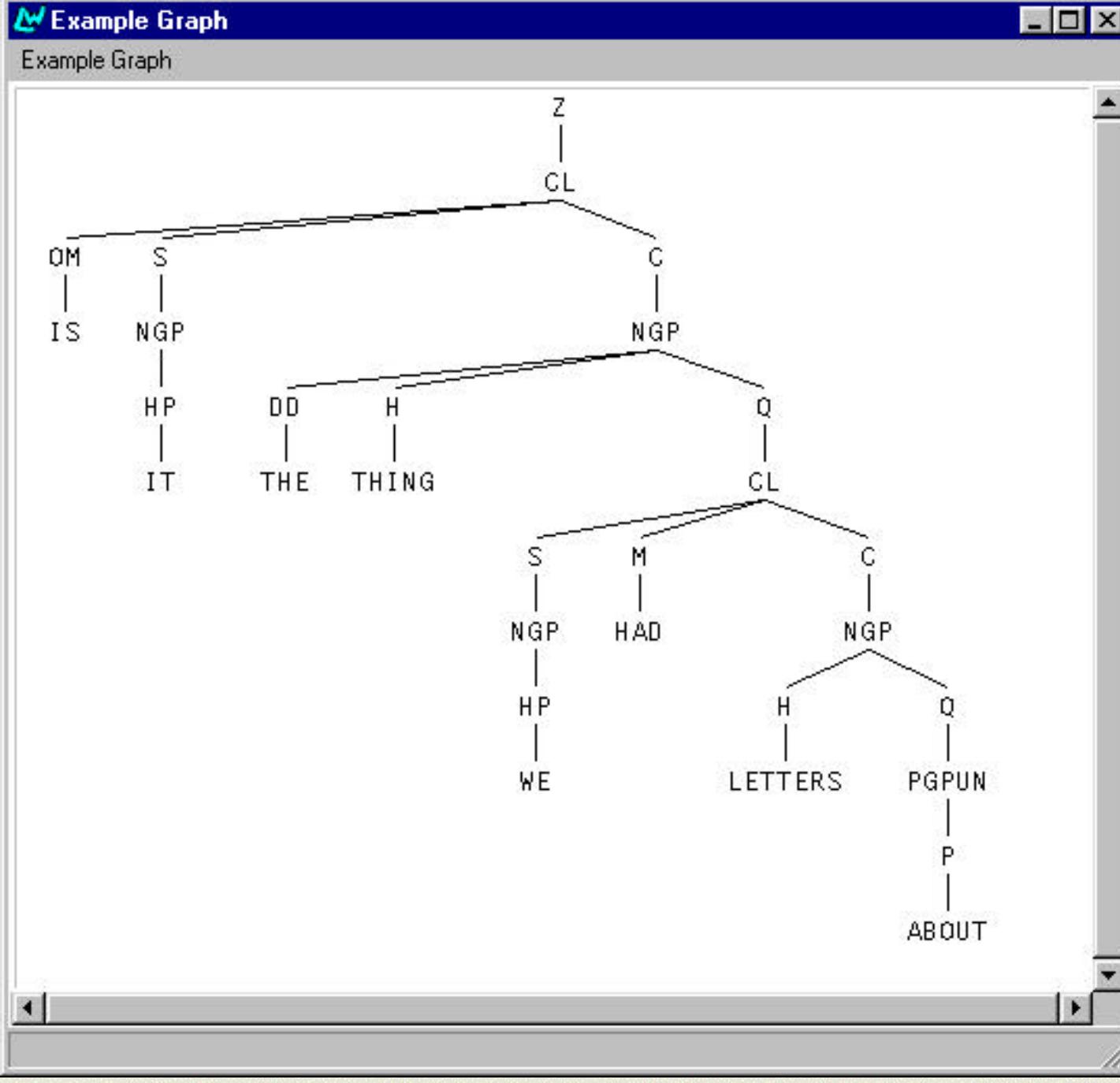


# Example entry on ICAME CD

Z 1 CL 2 S NGP HP I 2 M PLAY 2 C PGP 3 P WITH [FS:MY-  
CHIP] 3 CV NGP 4 DD MY [RP:MY] 4 MO QQGP AX BIG 4 H  
TIPPER-LORRY 1 CL 5 & AND 5 S NGP HP I [RP:I] 5 M  
CALL 5 C PGP 6 PM FOR 6 CV NGP HN DAVID

This can be converted to the XML form defined for systemic instantiated structures and then read into any tool that supports this. For example, KPML...





**Example Manager (KPML)**

Example Operation Commands File Edit Generate Mark Clear Modes Runner

CURRENT LANGUAGE IS: ENGLISH  
EXAMPLE SET SHOWN: 10BBPSHW (62 EXAMPLES)

marked	name	TARGET	generated? mismatch?
	10BBPSHW.3	WE COULD PUT THE FLOOR IN	•
	10BBPSHW.4	AND WE CAN MAKE FURNITURE	•
	10BBPSHW.7	(S) (OM) BE A GOOD IDEA THAT YES	•
	10BBPSHW.10	YES	•
	10BBPSHW.16	IT'S VERY NICE SET OF LEGO THIS	•
	10BBPSHW.21	IT'LL COLLAPSE	•
	10BBPSHW.23	IT'LL BE VERY GOOD THEN WON'T IT	•
	10BBPSHW.25	DRAT IT'S COLLAPSING	•
	10BBPSHW.30	WE CAN MAKE A CAT FOR THEM	•
	10BBPSHW.34	YEAH WE CAN MAKE A CAR TO GO IN THE...	•
	10BBPSHW.36	YOU MAKE THE CAR MARTYN	•
	10BBPSHW.39	ME AND GARETH WILL MAKE THE HOUSE A...	•
	10BBPSHW.41	YEAH FIRST OF-ALL WE'LL DO THE HOUS...	•
	10BBPSHW.46	(S) WONDER ID IT'S GOING-TO BE A GO...	•

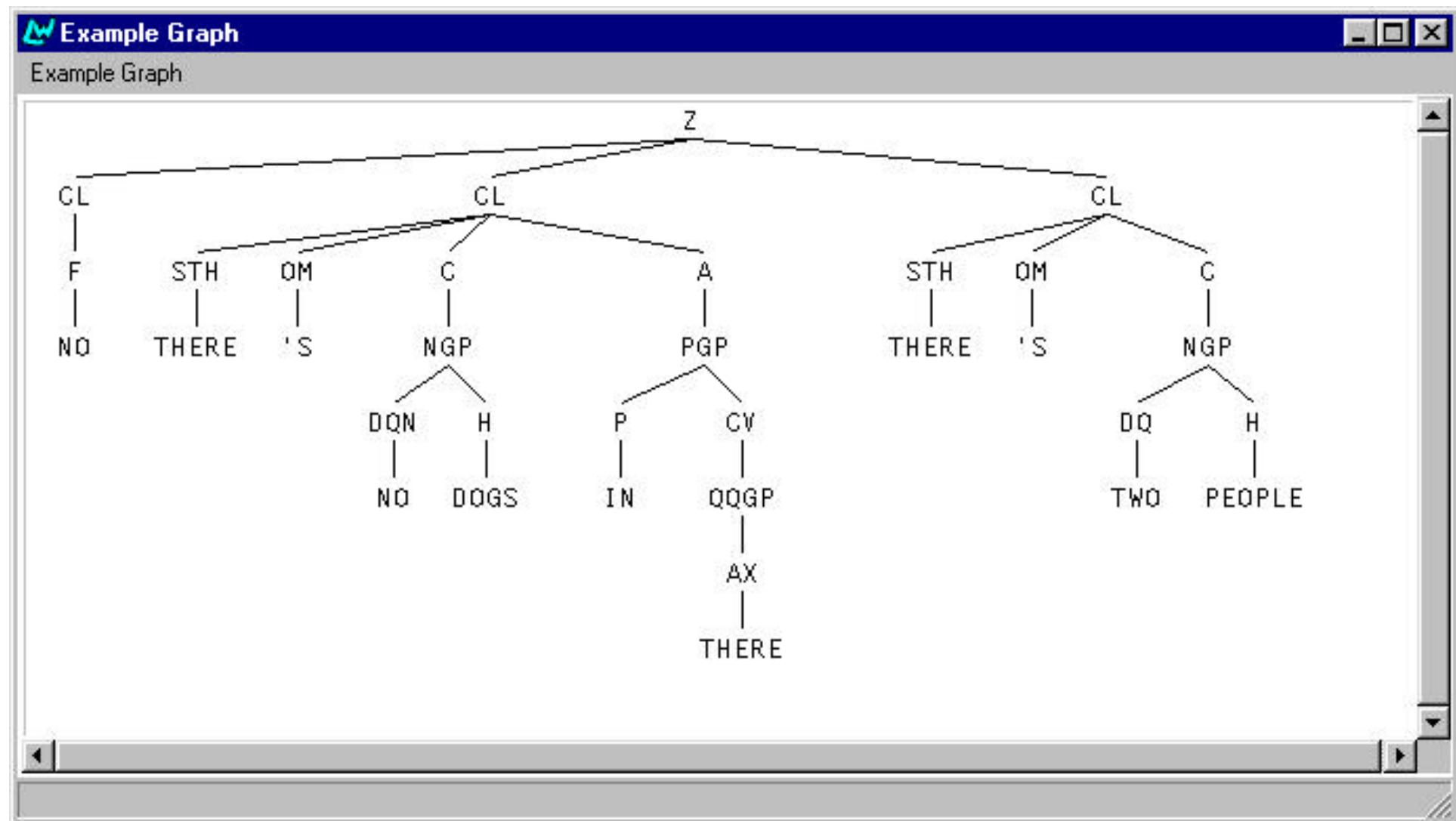
Example sets for ENGLISH loaded... 10ABIHS 10ABIRL 10ABITG 10ABPSHS 10ABPSRL 10ABPSTG 10AGIKP 10AGILB 10AGIRV 10AGPSKP 10AGPSLB 10AGPSRV 10BBIGJ 10BBIHW 10BBIMJ 10BBPSGJ 10BBPSHW 10BBPSMJ 10BGICL 10BGIEE 10BGIRE 10BGPSC 10BGPSEE 10BGPSPR E 10CBIAT 10CBILJ 10CBIMH 10CBPSAT 10CBPSLJ 10CBPSMH 10CGIED 10CGIJN 10CGISP 10CGPSED 10CGPSJN 10CGPSSP 10DBIDM 10DBP SDM 10DGIDH 10DGISM 10DGSS 10DGPSDH 10DGPSSM 151B 10DGPSS 12ABIW 12ABIPG 12ABISM 12ABPSAW 12ABPSPG 12ABPSSM 12AGIAH 12AGIBR 12AGIND 12AGPSAH 12AGPSBR 12AGPSND 12BBIGW 12BBIMB 12BBIMN 12BBPSGW 12BBPSMB 12BBPSMN 12BGINAH 12BGIH L 12BGISG 12BGPSSAH 12BGPSSHL 12BGPSSG 12CBIBW 12CBIML 12CBIMR 12CBPSBW 12CBPSML 12CBPSMR 12CGIJP 12CGISJ 12CGIWG 12CGPSJP 12CGPSSJ 12CGPSWG 35B 129B 12DBIAF 12DBILJ 20B 12DBIPL 12DBPSAF

PUNCTUATION: ON OFF CASE: ON OFF SORT: ON OFF GLOSSES: ON OFF DISPLAY: TARGET GENERATED NONE

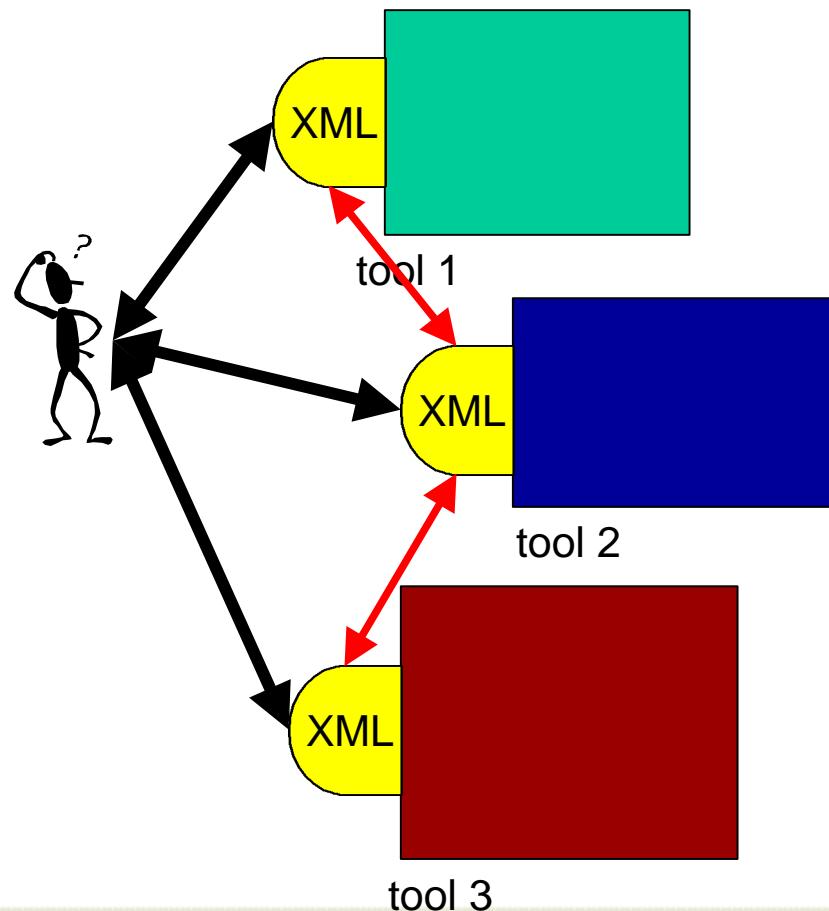
L: Restrict view to this example set; R: Menu of operations for this object.

- 10BBPSHW.191 THERE'S ANOTHER MAN -
- 10BBPSHW.192 THERE'S A MAN'S HEAD ●
- 10BBPSHW.193 THERE'S ANOTHER HEAD ●
- 10BBPSHW.194 NO THERE'S NO DOGS IN THERE THERE'S...** ●
- 10BBPSHW.198 VERY GOOD ●
- 10BBPSHW.202 YEAH ●
- 10BBPSHW.207 I THINK IT'S GOIN-TO TAKE A LONG TI... ●
- 10BBPSHW.210 (S) (OM) NOT (M) (C) (P) THE RATE W... ●
- 10RRPCHW.225 PARDON ●

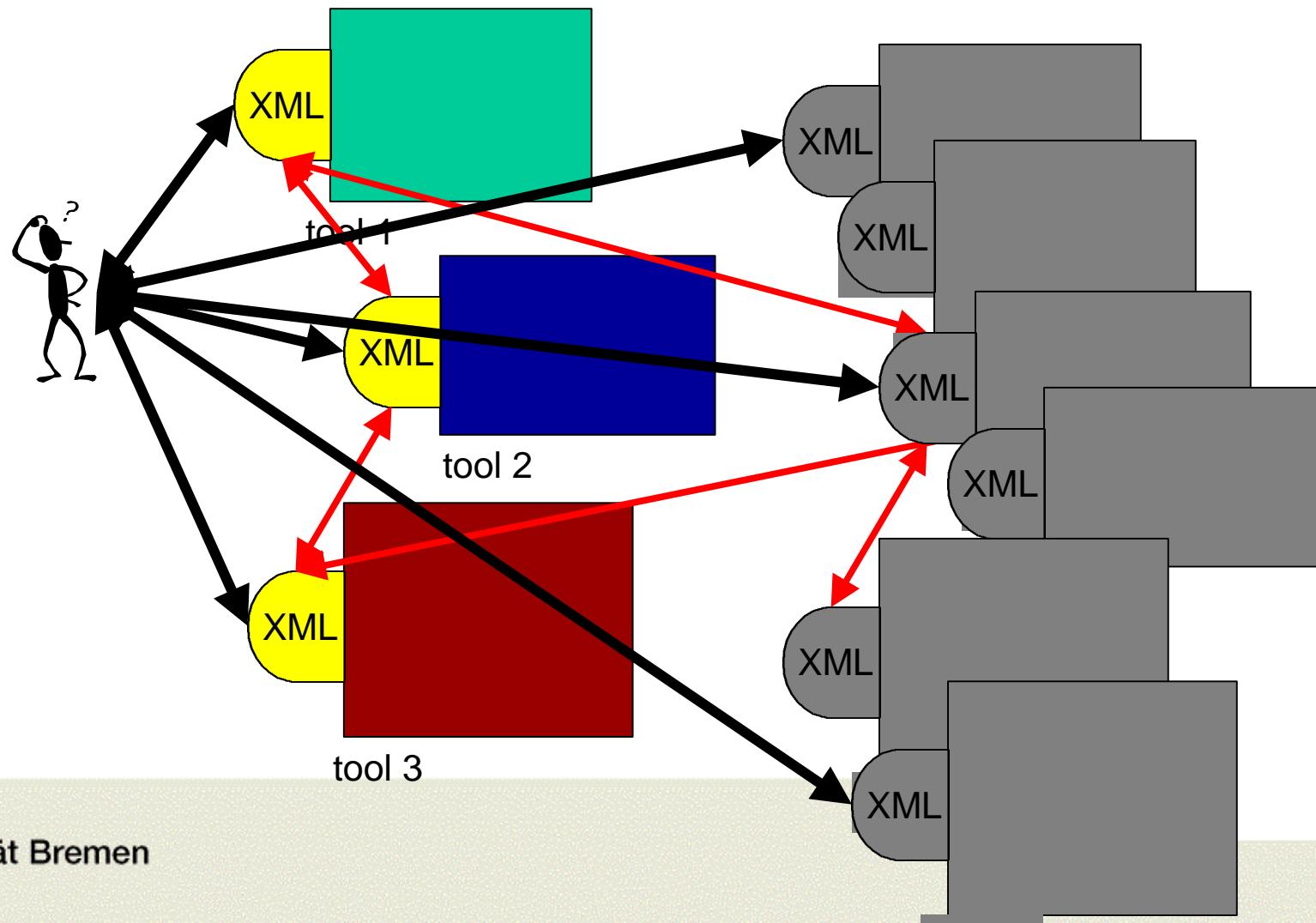




# 2000s view of linguistic tools



# 2000s view of linguistic tools



# Some boats to catch...

- where are the systemic-functional treebanks?
- if analyses are produced or ‘written up’ using a tool that can export XML structures for incorporation in a treebank, this would be of enormous value
  - supporting comparison of analyses
  - supporting collections of teaching examples
  - documenting the state of the art

